BLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

Robert FÉRON

INFORMATION, RÉGRESSION CORRÉLATION

VOL. V - FASCICULES 3-4 - 1956

PARIS
11, Rue Pierre Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse :

INSTITUT DE STATISTIQUE DE L'UNIVERSITE DE PARIS

INSTITUT HENRI POINCARÉ - 11, Rue Pierre Curie - PARIS (Ve)

Les manuscrits doivent être envoyés à *M. Daniel DUGUÉ*,
à l'adresse précédente.

Abonnements : Pour la France 1.700 francs français
Pour l'Etranger 2.000 francs

Vente au numéro : *(fascicule de 50 pages environ)*
Pour la France 500 francs
Pour l'Etranger 600 francs

UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

Robert FÉRON

INFORMATION, RÉGRESSION
CORRÉLATION

VOL. V - FASCICULES 3-4 - 1956

PARIS
11, Rue Pierre Curie

AVERTISSEMENT

Ce travail forme un tout cohérent et a été conçu de manière à pouvoir être lu de personnes ne possédant qu'une culture mathématique et probabiliste très rudimentaire. Nous n'avons pas recherché la plus grande généralité possible, mais la plus grande simplicité et la plus grande clarté.

Les références bibliographiques ne sont presque jamais nécessaires à l'intelligence même du texte. Le sujet traité étant entièrement vierge, aucune lecture préalable n'est indispensable. Nous avons délibérément laissé de côté les problèmes connexes des nôtres déjà partiellement résolus qui auraient nécessité un exposé trop long. Nous avons essayé de suppléer à cette carence en faisant suivre chaque chapitre et l'ensemble du travail d'une liste de lectures recommandées. Cette liste a été conçue comme un moyen de travail et de recherche. Nous nous sommes tout spécialement attachés à citer les auteurs qui apportaient des idées neuves. Aussi avons-nous été amenés à citer des auteurs qui ont fait un certain nombre d'erreurs, non que ces auteurs soient de mauvais mathématiciens, mais parce qu'ils traitaient des sujets très nouveaux et par conséquent très délicats et difficiles. Nous ne saurions donc trop recommander au lecteur de ces ouvrages de revérifier avec soin tous les calculs qu'il utilise. Le praticien pressé qui ne voudrait pas s'astreindre à ce travail fera bien de borner ses emprunts aux résultats énoncés dans le présent travail et à ceux des auteurs sûrs comme CRAMER, DARMOIS, FORTET, FRECHET.

L'ordre dans lequel nous avons exposé ces recherches n'est pas celui dans lequel nous les avons effectuées. En réalité, nous avions la plupart des résultats du chapitre IX avant ceux des autres chapitres. Le chapitre IX forme à lui seul un tout cohérent et peut sans inconvénient être lu le premier. Par contre, les autres chapitres doivent nécessairement être lus dans l'ordre où nous les avons placés.

Ces recherches entreprises dès 1947 n'ont pu être menées à bien que grâce à la bienveillance que mes directeurs de Recherches Maurice FRECHET et Georges DARMOIS m'ont toujours témoignée et aux encouragements qu'ils ont bien voulu me prodiguer lors de l'exécution de ce travail.

J'exprime ici toute ma reconnaissance, tout particulièrement au Professeur Maurice FRECHET pour les très nombreuses publications de mathématiciens éminents dont il m'a fait généreusement cadeau et pour le soin minutieux avec lequel il a relu tous mes travaux, et au Professeur

Georges DARMOIS qui ne s'est pas contenté de lire mon manuscrit, mais a lui-même travaillé dessus et auquel je dois au moins deux idées maitresses de cette thèse.

J'exprime aussi toute ma gratitude au Professeur E.S. PEARSON qui, durant toute une année, a spécialement consacré une heure par semaine de son précieux temps à ma formation particulière.

Je voudrais aussi remercier ici tous les professeurs dont j'ai suivi avec enthousiasme les cours de probabilité ou de statistique. Je ne citerai que les principaux : F.N. DAVID, P. DELAPORTE, J. DUBOURDIEU, D. DUGUE, R. FORTET, J.B.S. HALDANE, H.O. HARTLEY, R. HENON, M. JANET, N.L. JOHNSON, M.G. KENDALL, R. Von MISES, MORICE, J. NEYMAN, R. ROY, Miss THOMAS, N. WIENER.

Je voudrais enfin remercier les personnes avec lesquelles j'ai eu les conversations particulièrement fructueuses : ARNOUS, BARNARD, BARTLETT, R. BONNET, F. CANTELLI, H. CRAMER, VAN DANTZIG, J.L. DOOB, M-J. DUHAMEL, Mrs EVANS, FONTAPPIE, C. FOURGEAUD, A. FUCHS, C. GINI, HOSTINSKY, H. GEIRINGER, Th. GUILBAUD, J. JOUVIN, KAMPE DE FERIET, KARUHNEN, W. KOSTITZIN, LE CAM, P. LEVY, M. LOEVE, A. LUBIN, S. MARQUET, E. MOURIER, OTTAVIANI, PINEDA, F. POLLACZEK, POMPILJ, R. RAO, RIGUET, R. RISSER, S. RIOS, F. ROSENFELD, T. SALVEMINI, SAVAGE, M.P. SCHUTZENBERGER, C.A.B. SMITH, STEYLERS, P.V. SUKHATME, THIONNET, VAULOT, J. VILLE, J. WISHART, H. WOLD.

Mes remerciements aussi à tous ceux auxquels je ne pense pas en ce moment qui m'ont aidé dans la poursuite de mes recherches.

INTRODUCTION

Un des principaux buts de la statistique est la description d'une v.a. (en abrégé pour variable aléatoire) X de fonction de répartition (f. de r.) F (x) : au moyen d'un certain nombre de fonctionnelles de F (x) : ϕ [F(x)], que nous désignerons sous le nom de caractéristiques fonctionnelles (1).

Nous pouvons généraliser la notion de caractéristique fonctionnelle à l'espace à 2 dimensions. Etant donné un couple aléatoire XY de f. de r F (x, y), nous pourrons encore continuer à lui associer divers éléments de la forme ϕ [F(x,y)] , éléments que nous continuerons à désigner sous le nom de caractéristiques fonctionnelles (c.f.).

Certaines de ces c.f. sont des nombres comme par exemple :

- la moyenne de la v.a. Y
- la variance de la v.a. Y
- le coefficient de corrélation linéaire r du couple aléatoire XY.
- le rapport de corrélation η de Pearson.

D'autres sont des lignes comme par exemple :

- la ligne de régression des moyennes de Y en x.

Il est bien clair que chacune des 5 c.f. ci-dessus constitue pour nous un outil de travail et l'on peut citer maints problèmes qui peuvent être résolus grâce à leur emploi simultané. Ceci nous suggère qu'il doit exister entre eux un lien de parenté alors qu'il n'en existerait pas entre les 5 c.f. suivantes :

- la dominante de Y
- l'entropie de Y
- le coefficient de corrélation linéaire du couple XY.
- l'indice de connexion simple de Gini.
- la ligne de régression des médianes de Y en x.

Aussi avons-nous cherché à dégager le lien qui existe entre les 5 c.f.

Nous nous proposons dans cette introduction de donner quelques théorèmes simples dont le lecteur trouvera lui-même sans peine des

(1) On remarquera que nos caractéristiques fonctionnelles ne correspondent pas exactement à ce que les Anglo-Saxons désignent sous le terme de Statistics, les statistics étant en réalité des caractéristiques fonctionnelles empiriques.

démonstrations élémentaires. Ces théorèmes sont des cas particuliers de propositions beaucoup plus générales que nous établirons dans la suite et doivent dans notre esprit simplement aider le lecteur à suivre le fil directeur qui nous a guidés quand nous avons écrit ce travail. -

Bornons-nous tout d'abord au cas où X est une v.a. discrète (le seul où η^2 a été défini jusqu'à présent).

Avant nous, le cas unidimensionnel a été traité par Fréchet [2] (1) qui, dès 1946 a clairement mis en évidence qu'il existe une certaine parenté entre deux caractéristiques fonctionnelles particulières la variance et la moyenne (le même lien de parenté existant aussi par exemple entre l'écart moyen et la médiane).

Mais ce n'est qu'en 1949 que Georges Darmois [2] jeta quelque lumière sur la relation qui existe entre les autres notions.

Ce dernier remarque en effet que l'on peut se considérer comme d'autant mieux renseigné sur les valeurs que peut prendre une v.a. Y que la variance σ^2_Y de cette v.a. est plus petite. Il est clair que, si je sais en outre que la v.a. X a pris la modalité x_i, je serai amené à me considérer comme d'autant mieux renseigné que $\sigma^2_{Y_{x_i}}$ est plus petit. Georges Darmois remarque que l'on n'a pas toujours $\sigma^2_{Y_{x_i}} < \sigma^2_Y$ mais que par contre on a toujours :

$$\sum \sigma^2_{Y_{x_i}} \Pr(X = x_i) \leq \sigma^2_Y$$

et il traduit le fait que la quantité :

$$\Delta \mathfrak{I} = \sum (\sigma^2_Y - \sigma^2_{Y_{x_i}}) \Pr(X = x_i) \tag{2}$$

ne peut être négative en disant qu'un renseignement supplémentaire ne peut en moyenne nous faire perdre de l'information ; aussi considère-t-il la quantité $\Delta \mathfrak{I}$ comme un outil commode pour mesurer l'avantage que nous avons à connaitre une liaison.

Il est remarquable de constater que si nous formons le rapport $\frac{\Delta \mathfrak{I}}{\sigma^2_Y}$, nous obtenons le carré du rapport de corrélation de Pearson. η^2 apparait donc ainsi comme le rapport entre ce que nous avons effectivement gagné et le maximum de ce que nous pouvions espérer gagner. C'est là, pensons-nous, l'aspect le plus simple et le plus naturel de la notion de "corrélation".

Mais il existe à notre avis un autre aspect plus complexe, que nous désignerons sous le nom de "corrélation dure" parce que pour parvenir à cette notion, nous sommes amenés à considérer les surfaces de corrélation dure de S. Bernstein.

On peut en effet en première approximation remplacer l'étude du couple aléatoire XY par celle d'un couple aléatoire plus simple $X^* Y^*$ que nous appellerons <u>couple aléatoire associé au couple XY</u>.

Certes, il y a bien des manières de faire une telle substitution. Nous n'en considèrerons pourtant qu'une seule : celle où Y^* est en <u>corrélation dure</u> par rapport à x^* au sens de S. Bernstein [1] .

---

(1) Les chiffres entre crochets renvoient à la bibliographie placée à la fin du présent travail.

Généralisant légèrement les définitions de S. Bernstein, nous dirons que la v.a. Y est en corrélation dure relativement à x si les fonctions de répartition liées aux divers x_i sont identiques à une translation près.

Ceci nous amène à imposer au couple aléatoire $X^* Y^*$ associé au couple XY d'être tel que :

$$\Pr (Y^* < y) \mid X=x_i) = \Psi \left[y - \varphi (x_i) \right] \quad (3)$$

où

$$\Psi (y) = \sum_i F_{x_i} \left[y + \varphi (x_i) \right] \quad \Pr(X = x_i) \quad (4)$$

Ceci posé, il est bien clair qu'à toute courbe φ (x) les relations (3) et (4) font correspondre d'une manière unique un couple aléatoire $X^* Y^*$ associé au couple aléatoire XY.

On peut calculer le "gain d'information dure" $\Delta' \mathcal{J}_{X^* Y^*} = \sigma^2_Y - \sigma^2_{Y^*_{x_i}}$ relatif à chacun des couples aléatoires $X^* Y^*$ et on constate que $\Delta' \mathcal{J}_{X^* Y^*}$ est maximum quand on prend précisément :

$$\varphi (x_i) = \mathfrak{M} \; Y_{x_i} + K \quad (5)$$

et nous exprimons ce fait en disant que le schéma probabiliste est le mieux adapté possible quand φ (x) est une courbe parallèle à la ligne de régression des moyennes.

Mais il peut être aussi utile de considérer un autre schéma probabiliste moins bien adapté à la description du couple aléatoire X Y, mais qui présente sur le précédent l'avantage d'être considérablement plus simple. C'est celui où on impose aux φ (x_i) d'être d'une forme simple. Par exemple, on peut imposer aux points φ (x_i) d'être tous situés sur une droite. Quand cette dernière condition est vérifiée, on constate que le $\Delta' \mathcal{J}$ du couple $X^* Y^*$ est maximum précisément quand les points φ (x_i) se trouvent sur une droite parallèle à la droite ajustée par la méthode des moindres carrés. Si φ (x_i) est ainsi déterminé, on peut facilement montrer que le rapport entre l'information gagnée sur le couple $X^* Y^*$ et σ^2_Y (information maximum que nous pouvions espérer gagner) n'est autre que le carré du coefficient de corrélation linéaire r^2.

Dans le travail qui va suivre, nous nous proposons de généraliser les idées exprimées ci-dessus.

D'une part, nous ne nous bornerons plus à considérer l'imprécision de nos connaissances comme nécessairement déterminée par la variance, mais considèrerons une classe beaucoup plus vaste de c.f. (celle des fonctionnelles concaves). D'autre part, nous ne nous bornerons plus à imposer à X d'être une v.a. discrète, nous supposerons le couple aléatoire XY absolument quelconque.

CHAPITRE I

DE L'INCERTITUDE

INTRODUCTION

Depuis fort longtemps on a cherché à définir des indices pouvant servir à caractériser l'imprécision avec laquelle nous connaissons la valeur qu'est susceptible de prendre une v.a. (en abrégé pour variable aléatoire) Y de f.d.r. (en abrégé pour fonction de répartition) F (y).

Durant ces dernières années, on s'est plus particulièrement occupé de généraliser les indices connus.

C'est ainsi que l'écart moyen minimum (1), l'écart type et l'étendue sont apparus comme des écarts typiques de Fréchet d'ordre 1, 2, l'infini respectivement (cf. Fréchet [3]).

La notion de différence moyenne (2) a été reprise et généralisée (Brambilla [1]) ; on montre que cet indice peut se mettre sous la forme :

$$D = \int_{-\infty}^{+\infty} 2\, F(y) \left[1 - F(y) \right] dy$$

Enfin la notion d'entropie, introduite en téléphonie en 1928 par Hartley [1], s'est déjà montrée susceptible de généralisations.

1. — ESSAI DE DÉFINITION AXIOMATIQUE D'UN INDICE D'INCERTITUDE

Nous nous proposons ici de donner une définition axiomatique des indices précédents. Ces indices que nous appellerons indices d'incertitude seront des fonctionnelles de la fonction de répartition de Y :

$$\mathfrak{I} = \phi \left[F(y) \right]$$

et vérifieront certaines conditions que nous nous proposons de préciser.

---

(1) L'écart moyen est déjà comparé à l'écart type dans Helmert (1).

(2) Cette notion a été introduite semble-t-il par l'astronome allemand Jordan, à la fin du siècle dernier.

Parmi les conditions qu'on peut à priori songer à imposer à un indice d'incertitude, citons :

CONDITIONS QU'ON PEUT SONGER A IMPOSER :

condition I : Si l'incertitude $\mathfrak{J}_Y$ relative à la v.a. Y existe, l'incertitude $\mathfrak{J}_{\lambda Y}$ relative à la v.a. λY existe également et est telle que :

$$\mathfrak{J}_{\lambda Y} \geqslant \mathfrak{J}_Y \qquad \text{si} \quad \lambda > 1$$

autrement dit :

$$\phi\left[F(y)\right] \leqslant \phi\left[F\frac{y}{\lambda}\right] \qquad \text{si } \lambda > 1$$

condition II : Si l'incertitude relative à la v.a. Y existe, l' incertitude relative à la v.a. Y-a existe également et lui est égale, autrement dit :

$$\phi\left[F(y)\right] = \phi\left[F(y-a)\right]$$

condition III : L'incertitude relative au mélange de 2 populations également probables existe toutes les fois que l'incertitude relative à ces 2 populations existe et est supérieure ou égale à la moyenne de l'incertitude des 2 populations.

$$\phi\left[\frac{F_1(y) + F_2(y)}{2}\right] \geqslant \frac{1}{2}\left\{\phi\left[F_1(y)\right] + \phi\left[F_2(y)\right]\right\}$$

condition IV : L'incertitude relative au mélange de k populations dans des proportions p_1, p_2, ... p_k existe toutes les fois que l'incertitude relative à chacune de ces populations existe et est supérieure ou égale à la moyenne pondérée de l'incertitude des k populations, autrement dit :

$$\phi\left[\sum p_i F_i(y)\right] \geqslant \sum p_i \, \phi\left[F_i(y)\right]$$

condition V : L'incertitude est une quantité positive ou nulle et on a $\mathfrak{J}_Y = 0$ si Y est une v.a. presque certainement égale à une constante.

Cette liste n'est évidemment pas limitative ; on peut par exemple également songer à imposer à la fonctionnelle ϕ certaines conditions de continuité. D'autre part, il est clair que dans une définition axiomatique de l'incertitude, nous n'aurons pas besoin d'imposer toutes les conditions de I à V ; en effet, certaines de ces conditions sont plus ou moins les conséquences des autres ; certaines autres ne sont pas utiles pour la démonstration des théorèmes que nous avons en vue et il nous est apparu tout à fait inutile de restreindre ainsi la portée de ces théorèmes.

2. — LES INDICES CLASSIQUES VERIFIENT LES CONDITIONS I à IV

Nous allons voir que les conditions I à III sont bien vérifiées pour les indices classiques et comme il résultera des résultats du chapitre III que la condition IV l'est aussi, nous admettrons ce résultat sans démonstration.

Enfin, il est bien clair que la condition V est bien vérifiée si on prend pour mesure de l'incertitude le moment typique de Fréchet d'ordre k ou la différence moyenne. Cette condition continue d'être vérifiée si on suppose

que Y est une v.a. discrète et si on prend l'entropie pour mesure de l'incertitude. Elle cesse d'être vérifiée si on suppose que Y est une v.a. possédant une densité de probabilité f(y) et que $\phi\ [F\ (y)]$ est donné par :

$\phi\ [F(y)] = -\int f\ (y)\ L\ f(y)\ dy$, cette quantité pouvant évidemment être négative.

Montrons maintenant que les conditions de I à III sont vérifiées :

1) si nous prenons pour ϕ le moment typique de Fréchet d'ordre k

$$\phi\left[F(y)\right] = \min_a \int |y-a|^k \ d\,F(y)$$

α) $\phi\left[F\left(\frac{y}{\lambda}\right)\right]$ existe si $\phi\left[F(y)\right]$ existe et on a évidemment :

$$\begin{aligned}\phi\left[F\left(\frac{y}{\lambda}\right)\right] &= \min_b \int |y-b|^k \ d\,F\left(\frac{y}{\lambda}\right) \\ &= \min_{b'} \int |\lambda\xi - b'|^k \ d\,F\,(\xi) \\ &= \lambda^k \ \phi\left[F(y)\right]\end{aligned}$$

et on a bien :

$$\phi\left[F(y)\right] \leqslant \phi\left[F\left(\frac{y}{\lambda}\right)\right] \qquad \text{si} \quad \lambda > 1$$

β) si $\phi\left[F(y)\right]$ existe, il est clair que $\phi\left[F(y-a)\right]$ existe également et qu'on a :

$$\begin{aligned}\phi\left[F(y-a)\right] &= \min_b \int |y-b|^k \ d\,F\,(y-a) \\ &= \min_{b'} \int |z+a-b|^k \ d\,F\,(z) \\ & \phi\left[F\,(y)\right]\end{aligned}$$

γ) enfin la quantité :

$$\phi\left[\frac{F(y)+G(y)}{2}\right] = \min_b \int \left|y-b\right|^k \ d\,\frac{F(y)+G\,(y)}{2}$$

est évidemment définie si les moments d'ordre k de Y et Y' existent. Soit alors b^* une valeur de b pour laquelle le minimum de l'intégrale précédente est atteint. On aura évidemment :

$$\phi\left[\frac{F(y)+G(y)}{2}\right] = \frac{1}{2}\left\{\int |y-b^*|^k \ d\,F(y) + \int |y-b^*|^k \ d\,G(y)\right\}$$

et le second membre est évidemment supérieur ou égal à :

$$\frac{1}{2}\left\{\min_{b'} \int |y-b'|^k \ d\,F(y) + \min_{b''} \int |y-b''|^k \ d\,G(y)\right\}$$

donc :

$$\phi\left[\frac{F(y)+G(y)}{2}\right] \geqslant \frac{1}{2}\left\{\phi\left[F(y)\right]+\phi\left[G(y)\right]\right\}$$

2) Considérons maintenant le cas de la différence moyenne D.

On a :

$$\begin{aligned}\phi\left[F\left(\frac{y}{\lambda}\right)\right] &= 2\int F\left(\frac{y}{\lambda}\right)\left[1-F\left(\frac{y}{\lambda}\right)\right] dy\\ &= 2\int F(z)\left[1-F(z)\right]\lambda \quad dz\\ &= \lambda\,\phi\left[F(y)\right]\end{aligned}$$

donc :

$$\phi\left[F\left(\frac{y}{\lambda}\right)\right] > \phi\left[F(y)\right] \quad \text{si} \quad \lambda > 1$$

et la condition I est bien vérifiée.

β) On voit de même que :

$$\phi\left[F(y-a)\right] = \phi\left[F(y)\right]$$

γ) enfin, on a :

$$\begin{aligned}\phi\left[\frac{F(y)+G(y)}{2}\right] &= 2\int \frac{F(y)+G(y)}{2}\left[1-\frac{F(y)+G(y)}{2}\right] dy\\ &= \frac{1}{2}\left\{2\int F(y)\left[1-F(y)\right] dy + 2\int G(y)\left[1-G(y)\right] dy\right\}\\ &\qquad + \frac{1}{2}\int\left[F(y)-G(y)\right]^2 dy\\ &= \frac{1}{2}\left\{\phi\left[F(y)\right]+\phi\left[G(y)\right]\right\}+\frac{1}{2}\int\left[F(y)-G(y)\right]^2 dy\end{aligned}$$

et la condition III est bien vérifiée.

3) Enfin l'entropie n'est définie que dans 2 cas particuliers, le cas où Y est une v.a. aléatoire discrète et celui où Y possède une densité de probabilité.

α) <u>entropie dans le cas discret</u>.

On suppose que Y est susceptible de prendre un nombre fini de valeurs $y_1, y_2, \ldots y_n$ avec des probabilités $p_1, \ldots p_n$

On aura alors :

$$\phi\left[F(y)\right] = -\sum p_i \, L \, p_i$$

ou en posant :

$$-p\,L\,p = H(p)$$

$$\phi\left[F(y)\right] = \sum H(p_i)$$

et il en résulte immédiatement que :

$$\phi\left[F\left(\frac{y}{\lambda}\right)\right] = \phi\left[F(y)\right]$$

$$\phi\left[F(y-a)\right] = \phi\left[F(y)\right]$$

De plus, si nous supposons que la v.a. Y' de fonction de répartition G(y) ne peut prendre que les valeurs $y_1, \ldots y_n$ avec les probabilités $p'_1, \ldots p'_n$, on aura :

$$\phi \left[\frac{F(y) + G(y)}{2} \right] = \sum H \left[\frac{(p_i + p_i')}{2} \right]$$

p_i et p_i' étant des quantités nécessairement positives ou nulles $H \left[\frac{(p_i + p_i')}{2} \right]$ sera nécessairement définie.

De plus, comme H est une fonction concave, on aura nécessairement :

$$\sum_i H \left[\frac{p_i + p_i'}{2} \right] \geqslant \sum_i \frac{1}{2} \left[H(p_i) + H(p_i') \right]$$

ce qui prouve la condition III.

β) <u>entropie dans le cas d'une densité de probabilité</u>.

On aura alors :

$$\phi \left[F(y) \right] = - \int f(y) \, L \, f(y) \, dy$$

$$\phi \left[F\left(\frac{y}{\lambda}\right) \right] = - \int \frac{1}{\lambda} \, f\left(\frac{y}{\lambda}\right) \left[L \, f\left(\frac{y}{\lambda}\right) - L\lambda \right] dy$$

$$= L\lambda + \phi \left[F(y) \right]$$

donc la condition I est bien vérifiée.

La condition II l'est aussi également.

Enfin, si $\int f(y) \, L \, f(y) \, dy$ et $\int g(y) \, L \, g(y) \, dy$ existent au sens classique. Alors :

$$\phi \left[\frac{F(y) + G(y)}{2} \right] = - \int \frac{f(y) + g(y)}{2} \, L \, \frac{f(y) + g(y)}{2} \, dy$$

existera aussi et sera telle que :

$$\phi \left[\frac{F(y) + G(y)}{2} \right] \geqslant \frac{1}{2} \left\{ \phi \left[F(y) \right] + \phi \left[G(y) \right] \right\}$$

3. — NOTRE DÉFINITION DE LA NOTION D'INCERTITUDE

Il résulte des paragraphes précédents que diverses définitions de la notion d'incertitude sont également possibles. Toutefois, comme le reste de notre travail roule sur l'étude de certaines propriétés de cet indice, nous aurons besoin d'adopter une définition précise de cette notion.

Il est clair que si nous adoptons une définition trop vague, les théorèmes auxquels nous parviendrons dans la suite seront d'un énoncé relativement fort compliqué. Au contraire, si nous adoptons une définition trop stricte, ces théorèmes seront d'un énoncé très court, mais perdront beaucoup de leur généralité.

Nous avons choisi un moyen terme entre ces deux écueils et dirons dans la suite que $\mathfrak{J}_Y$ est un indice d'incertitude s'il vérifie les conditions II et III.

<u>Définition</u>.

<u>Nous dirons donc que</u> $\phi \left[F(y) \right]$ <u>est un indice d'incertitude</u> si :

1° - <u>toutes les fois que</u> $\phi \left[F(y) \right]$ et $\phi \left[G(y) \right]$ <u>sont définis</u> $\phi \left[F(y-a) \right]$ <u>et</u>

$$\phi \left[\frac{F(y)+G(y)}{2}\right]$$ le sont aussi.

2° - ϕ est une "fonctionnelle concave", autrement dit :

$$\phi \left[\frac{F(y)+G(y)}{2}\right] \geqslant \frac{1}{2} \left\{ \phi \left[F(y)\right] + \phi \left[G(y)\right] \right\} \qquad (1.1)$$

3° - On a toujours $$\phi \left[F(y-a)\right] = \phi \left[F(y)\right] \qquad (1.2)$$

On verra dans la suite que dans notre énoncé, la condition de concavité (1.1) est une condition essentielle alors que la condition (1.2) est une condition accessoire qui permet simplement d'alléger l'énoncé de certains théorèmes.

Remarque : Nous avons préféré dans le cours de cet exposé éviter le mot d'information autant que faire se pouvait attendu que ce mot a été déjà employé dans des sens différents par Wiener (1) et Guilbaud (1) d'une part, Mandelbrot (1) et Schützenberger (1) de l'autre. - La meilleure de ces définitions et la seule employée au moment où nous avons écrit ce travail est sans contredit celle de Wiener qui appelle information une entropie négative $\mathfrak{I} = \Sigma\, p_i\, L\, p_i$ dans le cas discret. $\mathfrak{I} = \int f(y)\, L\, f(y)\, dy$ dans le "cas continu" - Cette définition est justifiée par le fait que l'information de Wiener augmente si la variable aléatoire Y est connue avec une plus grande précision. - Aussi nous conformant à l'idée directrice de Wiener, nous appellerons Information une incertitude négative. - Cette définition nous sera utile quand, au chapitre III, nous définirons notre gain d'information qui généralise le gain d'information de Wiener (1).

4. — DOMAINE DE DÉFINITION

Il est à noter que $\phi \left[F(y)\right]$ n'est pas en général définie en tout point de l'espace E des fonctions de répartition. Ceci nous conduit à associer à toute mesure de l'incertitude un domaine D de E qui devra être tel que :

1° si $F(y) \in D$ alors $\phi \left[F(y)\right]$ est définie

2° si $F(y) \in D$ et $G(y) \in D$ alors $\frac{F(y)+G(y)}{2} \in D$

les domaines D qui vérifient les conditions 1° et 2° jouissent d'importantes propriétés, aussi les désignerons-nous sous le nom de domaines de définition de ϕ.

D'autre part, à l'ensemble des v.a. $Y^{(r)}$ dont le moment absolu d'ordre r

$$m^{(r)} = \int_{-\infty}^{+\infty} |y|^r \, dF(y) \qquad (1.3)$$

existe, nous pouvons associer l'ensemble $\Delta^{(r)}$ des points F(y) de E tels que (1.3) ait un sens.

Il est alors clair que si nous prenons le moment typique d'ordre k de Fréchet comme mesure de l'incertitude, nous pourrons prendre le domaine $\Delta^{(k)}$ pour domaine de définition de ϕ.

(1) En théorie des télécommunications on utilise généralement le logarithme de base 2 - Dans notre théorie au contraire le choix de la base des logarithmes est sans importance.

De plus, il est clair que $\phi\left[F(y)\right]$ n'est pas défini si F(y) n'est pas un élément de $\Delta^{(k)}$. Nous traduirons ce fait en disant que $\Delta^{(k)}$ est le domaine maximum de définition de ϕ. Nous pourrons dans certains cas avoir intérêt à considérer certains domaines de définition de ϕ qui ne coincident pas avec le domaine maximum de définition de ϕ.

Dans le cas présent, nous pourrons par exemple prendre comme domaine de définition de ϕ tout domaine $\Delta^{(r)}$ où $r \geqslant k$.

De même, on ne peut prendre E comme domaine de définition de la différence moyenne ; toutefois il est possible de montrer que l'on peut prendre le domaine $\Delta^{(1)}$ défini plus haut comme domaine de définition de cette quantité.

De même encore pour l'entropie, on ne pourra prendre E comme domaine de définition.

Aussi nous sommes-nous demandé s'il existe des indices d'incertitude dont E soit précisément le domaine maximum de définition.

La réponse à cette question est affirmative.

Si nous prenons pour indice d'incertitude la valeur pour ℓ fixé de la fonction de concentration de Paul Lévy, changée de signe,

$$C = \phi\left[F(y)\right] = \min_{y} \left\{ F(y) - F(y+\ell) \right\}$$

Il résulte d'un théorème de G. Darmois [1] que la condition III est bien vérifiée. Il en est de même évidemment de la condition II. C'est donc bien un indice d'incertitude qui reste bien défini quel que soit F(y).

5. — DE L'UTILISATION DES INDICES D'INCERTITUDE

Il est clair qu'on ne pourra jamais connaitre l'indice d'incertitude théorique exactement. En revanche, on peut espérer connaitre l'indice d'incertitude approximativement en considérant les valeurs $y_1, y_2, \dots y_i, \dots y_n$, prises par une v.a. Y de fonctions de répartition F(y), au cours de n épreuves indépendantes.

On sait, d'après le théorème de Glivenko-Cantelli, que si n tend vers l'infini, la fonction de répartition empirique $F_n(y)$ converge presque sûrement vers F(y) et ceci uniformément sur tout l'intervalle $(-\infty, +\infty)$.

Définition.

Nous dirons avec Fréchet (cf. bibl. Fréchet [1]) que ϕ est une fonctionnelle continue en F(y) sur un domaine D de définition de ϕ si, quelle que soit la fonction G(y) de D et le nombre positif η, on peut trouver ε suffisamment petit pour que :

$$\operatorname{Max}\left| F(y) - G(y)\right| < \varepsilon \quad \text{entraine} \quad \left| \phi\left[F(y)\right] - \phi\left[G(y)\right]\right| < \eta$$

Dès lors, si ϕ est une fonctionnelle continue en F(y) sur D et si F (y) est un élément de D, alors d'après le théorème de Glivenko-Cantelli $\phi\left[F(y)\right]$ tendra presque sûrement vers $\phi\left[F(y)\right]$ quand n tend vers l'infini.

Le désir où nous sommes d'appliquer le théorème de Glivenko-Cantelli nous conduira souvent à ne pas prendre pour D le domaine maximum de définition de ϕ.

Ainsi, si nous prenons l'entropie pour mesure de l'incertitude, nous constatons que $\phi\,[F_n(y)]$ tendra vers $\phi\,[F(y)]$ si nous prenons pour D l'ensemble des F(y) totalement discontinues.

De même, si nous prenons la différence moyenne de Gini ou le moment d'ordre k de Fréchet $\phi\,[F_n(y)]$ tendra vers $\phi\,[F(y)]$ - si nous prenons pour D l'ensemble des fonctions F(y) correspondant à des variables aléatoires d'étendue finie.

Plus précisément, on peut montrer que dans le cas où on prend le moment d'ordre k de Fréchet comme mesure de l'incertitude, il n'est pas nécessaire d'imposer à F(y) d'être un élément de tous les $\Delta^{(r)}$ (définis p. 1.9), mais qu'il suffit pour que $\phi\,[F_n(y)] \longrightarrow \phi\,[F(y)]$ de prendre simplement $D = \Delta^{(2k)}$

LECTURES RECOMMANDEES

Les personnes désireuses de se documenter plus complètement sur quelques questions abordées ici liront avec fruit les articles suivants :

a) pour les valeurs typiques Fréchet [3] [7] [8] [15] Féron [2] [8] Jackson [1] Ville [1]

b) pour la différence moyenne Cantelli [1] De Finetti [1] De Finetti Piacello [1] Kendall [1] Salvemini [1]

c) pour l'entropie l'ouvrage fondamental reste Shannon et Weaver [1] on peut lire aussi Féron et Fourgeaud [1] C. Guilbaud [1] Wiener [1] et [2] Schützenberger [1] Mandelbrot [1] Hartley[1].

CHAPITRE II

DE LA FONCTION DE RÉPARTITION LIÉE

1. — RAPPEL DE DÉFINITIONS CLASSIQUES

Comme nous l'avons dit dans notre introduction, le rapport de corrélation de Pearson n'a été à notre connaissance défini que dans le cas où X est une v.a. discrète et il en va de même des principaux indices dits de corrélation. La raison de ce manque manifeste de généralité nous semble due au fait que ces indices ne peuvent être construits facilement que si on sait ce que l'on doit entendre par le mot "fonction de répartition liée".

Or, si cette notion est facile à définir dans le cas où X est une v.a. discrète, elle l'est si peu dans le cas général, qu'il nous eut été impossible de démontrer des théorèmes d'existence si la théorie de la mesure et des fondements du calcul des probabilités n'avaient été préalablement suffisamment développés.

Comme la plupart des auteurs, nous considérons que l'on peut associer à tout intervalle I de l'espace euclidien à 2 dimensions R_2 et plus généralement à tout ensemble de Borel S de R_2 une fonction P (S) appelée probabilité de S et vérifiant les conditions suivantes :

$$\left.\begin{array}{ll} 1°/ & P(S) \text{ est non négative } \quad P(S) \geqslant 0 \\ 2°/ & P(S) \text{ est complètement additive} \\ & P(S_1)+P(S_2)+\ldots+P(S_n)+\ldots = P(S_1+S_2+\ldots+S_n+\ldots) \\ & \text{si les } S_i \text{ sont disjoints} \\ 3°/ & P(R_2) = 1 \end{array}\right\} \quad (2.1)$$

Par un léger abus de langage, nous désignerons par $Pr(x_1 < X < x_2,\ y_1 < Y < y_2)$ la valeur de la fonction P(S) quand S est l'intervalle ouvert défini par les points $M_1(x_1, y_1)$ $M_2(x_2, y_2)$ et par $Pr\{X \subset T,\ y_1 < Y < y_2\}$ la valeur de P(S) quand S désigne l'ensemble des points x, y tels que $y_1 < Y < y_2$ et X est un point de l'ensemble T de l'espace R_1 (T est supposé B-mesurable).

Nous noterons simplement par $Pr(X \subset T)$ où T désigne un ensemble de points de l'espace R_1 la valeur de P(S) quand S désigne l'ensemble des points x, y, tels que X soit un élément de T (y restant quelconque).

Ces définitions classiques étant données, il est important de remarquer que la probabilité liée au sens classique que nous noterons

$Pr(y_1 < Y < y_2 \mid X \subset T)$ est définie seulement si $Pr(X \subset T)$ est définie et non nulle et dans ce cas, on a :

$$Pr(y_1 < Y < y_2 \mid X \subset T) = \frac{Pr[y_1 < Y < y_2, X \subset T]}{Pr(X \subset T)} \qquad (2.2)$$

et la fonction de répartition liée correspondante sera :

$$F(y \mid X \subset T) = Pr(Y \leqslant y \mid X \subset T) \qquad (2.3)$$

Il est à remarquer que $F(y \mid X \subset T)$ n'est définie au sens classique que si $Pr(X \subset T)$ est définie et positive.

En particulier $F(y \mid X=x)$ ne sera définie au classique qu'aux points x_i où $Pr(X = x_i) \neq 0$.

Il est clair que toute tentative se proposant d'étendre la notion classique de fonction de répartition liée présentera une certaine part d'arbitraire, afin d'étendre la notion de probabilité liée de la manière la plus utile, il convient tout d'abord d'étudier les propriétés de ces fonctions.

2. — SYSTÈME DE FONCTIONS DE RÉPARTITIONS LIÉES GÉNÉRALISÉES

1°) Si $T_1, T_2, \ldots$ sont disjoints, on a évidemment :

$$\sum_{i=1}^{\infty} F(y \mid T_i)\, Pr(X \subset T_i) = Pr\left\{X \subset T_1+T_2+\ldots+T_n+\ldots, y\right\} \qquad (2.4)$$

Ceci nous conduit à imposer à tout ensemble de fonctions de répartition liées généralisées $F_x(y)$ d'être tel que pour tout ensemble E de R_1 :

$$\int_E F_x(y)\, d\,P(T) = Pr(X \subset E,\ Y \leqslant y) \qquad (2.5)$$

L'intégrale étant bien entendu prise au sens de Lebesgue-Stieltjes; en langage probabiliste ordinaire (2.5) est complètement équivalente à :

$$\int_{-\infty}^{x} F_\xi(y)\, d\,F(\xi, \infty) = F(x,y) \qquad (2.6)$$

si $F(x,y) = Pr(X \leqslant x,\ Y \leqslant y)$.

2°) Nous verrons que si $\frac{\partial^2 F(x,y)}{\partial x\, \partial y}$ existe et est continue en tout point par rapport à l'ensemble des variables x, y, la ligne de régression des moyennes de y en x $y = \varphi(x)$ est telle que :

$$\varphi(x) = \lim_{T_i \to 0} \int_{-\infty}^{+\infty} y\, d\,F(y \mid T_i) \qquad (2.7)$$

Quand T_i est un intervalle ouvert contenant le point x. Comme nous désirons que les propriétés connues des v.a. possédant une densité de probabilité subsistent dans notre théorie, nous serons naturellement amenés à supposer que, si nous étudions les moyennes liées, on a :

$$\int y\, d\,F_x(y) = \lim_{T_i \to 0} \int y\, d\,F(y \mid T_i)$$

plus généralement si nous étudions la fonctionnelle $\phi\left[F(y)\right]$ (par exemple la variance liée, l'étendue liée, etc...) il nous sera commode d'imposer à $F(y)$ d'être tel que :

$$\phi\left[F_x(y)\right] = \liminf_{T_i \to O} \phi\left[F(y) \mid T_i)\right] \tag{2.8}$$

quand T_i est un intervalle ouvert contenant le point x.

3. — ENSEMBLES ADMISSIBLES DE FONCTIONS DE RÉPARTITION LIÉES

On remarquera que les conditions précédentes ne sont pas suffisantes pour nous permettre de définir $F_x(y)$ d'une manière unique en tout point du plan. Mais elles nous permettent de définir des ensembles E de fonctions de répartition qui se comporteront d'une manière identique dans nos calculs. De tels ensembles seront dits ensembles admissibles de fonctions de répartition si seule (2.5) est vérifiée, ensembles admissibles relativement à la fonctionnelle ϕ, si (2.5) et (2.8) le sont. Plus précisément :

Définition I.

Nous dirons que l'ensemble E des fonctions de répartition $F_x(y)$ est admissible si $\int_{-\infty}^{x} F_\xi(y)\, d\, F(\xi, \infty)$ existe et est égal à $F(x,y)$ l'intégrale étant prise au sens de Lebesgue-Stieltjes.

Il est à remarquer que cette intégrale existera pour tout ensemble de fonctions $F_x(y_0) = \gamma(x, y_0)$ mesurables.

En particulier, comme toute fonction mesurable au sens de Borel l'est au sens de Carathéodory, l'intégrale existera toujours si $\gamma(x, y_0)$ est B-mesurable. D'une manière plus particulière encore toute fonction continue étant B-mesurable, l'intégrale existera toujours si $\gamma(x, y_0)$ est continue.

Définition II.

Nous dirons que l'ensemble E des fonctions de répartition $F_x(y)$ est admissible relativement à la fonctionnelle ϕ, si $F_x(y)$ vérifie les conditions de la définition I et si de plus, quand $\liminf_{T_i \to O} \phi\left[F(y \mid T_i)\right]$ existe, quand T_i est un intervalle ouvert contenant le point x, alors $\phi\left[F_x(y)\right]$ existe et est tel que :

$$\phi\left[F_x(y)\right] = \liminf_{T_i \to O} \phi\left[F(y \mid T_i)\right] \tag{2.9}$$

Ceci posé, remarquons dès à présent que le système de fonctions de répartition liées, considéré par Paul Lévy (1) n'est pas nécessairement un système de fonctions de répartition admissible à notre sens. En effet, à ce stade de la théorie, la fonction de répartition est pour Paul Levy, une fonction croissante et bornée, mais on n'a pas nécessairement pour lui

$$\lim_{y \to -\infty} F_x(y) = 0 \qquad \lim_{y \to +\infty} F_x(y) = 1 \tag{2.10}$$

(1) Paul Lévy [1] p. 68.

Ces conditions sont pour nous, au contraire, absolument essentielles.

La même remarque s'applique à la définition de Madame Gruzewska 1 qui appelle fonction de répartition conditionnelle, la quantité :

$$\mathcal{G}(y) = \limsup_{\Delta x \to 0} \frac{F(x+\Delta x, y) - F(x,y)}{F(x+\Delta x, \infty) - F(x, \infty)} \tag{2.11}$$

et elle démontre le théorème suivant :

<u>Théorème de Mme Gruzewska</u> :

Si quand Δx tend vers 0, la quantité :

$$\delta(x,y,\Delta x) = \frac{F(x+\Delta x, y) - F(x,y)}{F(x+\Delta x, \infty) - F(x, \infty)} \tag{2.12}$$

tend vers une limite unique sur tout l'ensemble de définition (c'est-à-dire en tous les points tels que :

$$F(x+\Delta x, \infty) > F(x, \infty) \text{ pour tout } \Delta x > 0).$$

Alors on a :

$$F(x,y) = \int_{-\infty}^{x} \mathcal{G}_{\xi}(y)\, dF(\xi, \infty)$$

Ce théorème nous est apparu comme un cas particulier du théorème plus général suivant :

<u>Théorème I.</u>

<u>Quelle que soit la fonction de répartition F(x, y), il existe toujours une infinité de systèmes admissibles de fonctions de répartition liées.</u>

Ce théorème est un cas particulier d'un théorème beaucoup plus général dû à Doob (2) dont nous ne pouvons ici donner qu'une vague idée (cf. Doob [1] p. 18).

Etant donné un certain espace abstrait Ω et un certain corps de Borel $\mathcal{F}$, Doob appelle probabilité conditionnelle d'un certain ensemble mesurable M relativement au corps $\mathcal{F}$ toute fonction $P(M \mid F)$ de 2 variables (l'ensemble M et le point ω) telle que pour tout $\Lambda \in \mathcal{F}$

$$\int_{\Lambda} P(M \mid \mathcal{F})\, dP = P(\Lambda\ M)$$

et démontre l'existence d'une infinité de "versions" de $P(M \mid \mathcal{F})$.

Dès lors, si nous considérons 2 fonctions B-mesurables $x(\omega)$, $y(\omega)$ et si nous prenons :

1°/ pour M l'ensemble des points de Ω pour lesquels $y(\omega) < y$.

2°/ pour $\mathcal{F}$ le corps de Borel $B(\mathcal{F}_0)$ engendré par la classe $\mathcal{F}_0$ des ensembles de points ω tels que $x(\omega)$ appartienne à un intervalle.

On peut montrer aisément qu'il existe une version $z(\omega)$ de $P(M \mid \mathcal{F})$ telle que $z(\omega') = z(\omega'')$ toutes les fois que $x(\omega') = x(\omega'')$, autrement dit,

---

(1) Melicer-Gruzewska [1] Sur la distribuante de deux variables aléatoires dépendantes C.R. 1951 p. 1256.

(2) Doob [2] avait donné précédemment une autre définition de la probabilité conditionnelle qui a été critiquée par Dieudonné [1] [2] [3]

telle que $P(M \mid \mathcal{F})$ ne dépende que de x. Une telle "version" des $P(M \mid \mathcal{F})$ sera un ensemble de fonctions de répartition liées admissible à notre sens.

En raison de son importance capitale, nous donnerons ici une démonstration élémentaire du théorème I.

Il nous suffira évidemment pour démontrer le théorème 1 de prouver qu'il existe un ensemble admissible $\{F_x(y)\}$ de fonctions de répartition liées puisque si $\{H_x(y)\}$ est un ensemble de fonctions de répartition telles que $H_x(y) = F_x(y)$ sauf peut-être pour un ensemble de valeurs de x de probabilité nulle, on aura d'après la définition de l'intégrale de Lebesgue-Stieltjes :

$$F(x,y) = \int_{-\infty}^{x} F_{\xi}(y)\, dF(\xi, \infty)$$

$$= \int_{-\infty}^{x} H_{\xi}(y)\, dF(\xi, \infty) \qquad (2.13)$$

Pour montrer qu'il existe effectivement un système admissible de fonctions de répartition liées, remarquons que $F(x, \infty) = A(x)$ a, au plus, une infinité dénombrable de points de discontinuité. On pourra donc écrire :

$$A(x) = A_1(x) + \mathcal{A}(x)$$

où $A_1(x)$ est une fonction totalement discontinue, non-négative et non-décroissante,
et où $\mathcal{A}(x)$ est une fonction continue non-négative et non décroissante; soit S_1 l'ensemble des points où $A_1(x)$ présente des discontinuités. Si x_i est un point de l'ensemble S_1, on pourra poser :

$$F^*_{x_i}(y) = \frac{F(x_i, y) - F(x_i - 0, y)}{A(x_i) - A(x_i - 0)} \qquad (2.14)$$

Le dénominateur étant toujours positif, $F^*_{x_i}(y)$ sera toujours définie.

De plus, en tenant compte du fait que :

$$\lim.\ P(S_n) = P(\lim.\ S_n)$$

on voit en prenant successivement pour S_n l'ensemble des points tels que :

$$y \leqslant -n, x = x_i \text{ et } y \leqslant n,\ x = x_i\ n = 1, 2, 3\ldots$$

que :

$$\lim_{y \to -\infty} F^*_{x_i}(y) = 0$$

$$\lim_{y \to +\infty} F^*_{x_i}(y) = 1$$

posons par ailleurs $F_1(x,y) = \sum_{x_i < x} F(x_i, y) - F(x_i - 0, y)$

On aura évidemment :

$$F_1(x,y) = \int_{-\infty}^{x} F^*_{\xi}(y)\, dA_1(\xi)$$

posons $$\mathcal{F}(x,y) = F(x,y) - F_1(x,y) \qquad (2.15)$$

$\mathcal{F}(x,y)$ sera évidemment une fonction continue de x puisque si $x_2 > x_1$:

$$\mathcal{F}(x_2, y) - \mathcal{F}(x_1, y) \leqslant \mathcal{A}(x_2) - \mathcal{A}(x_1) \qquad (2.16)$$

et comme $\mathcal{A}(x)$, elle sera non-décroissante et bornée. Donc $\mathcal{A}(x)$ et $\mathcal{F}(x,y)$ seront des fonctions de x uniformément continues (1) et on voit comme précédemment que :

$$\lim_{y \to -\infty} \mathcal{F}(x_2,y) - \mathcal{F}(x_1,y) = 0$$

$$\lim_{y \to +\infty} \mathcal{F}(x_2,y) - \mathcal{F}(x_1,y) = \mathcal{A}(x_2) - \mathcal{A}(x_1)$$

Dès lors, posons $\mathcal{A}(x) = u$ ou $x = \alpha(u)$ et considérons la fonction $\mathcal{F}(\alpha(u), y_o) = \mathcal{F}_1(u, y_o)$. Cette fonction sera également continue et croissante et on a :

$$0 \leqslant \mathcal{F}_1(u, y_o) \leqslant 1$$

De plus, en vertu de (2.15), on aura :

$$\mathcal{F}_1(u+h, y_o) - \mathcal{F}_1(u, y_o) \leqslant h$$

$$\lim_{y \to -\infty} \mathcal{F}_1(u+h,y) - \mathcal{F}_1(u,y) = 0$$

$$\lim_{y \to +\infty} \mathcal{F}_1(u+h,y) - \mathcal{F}_1(u,y) = h$$

$F_1(u, y_o)$ ne sera donc pas seulement une fonction continue elle sera encore absolument continue et on aura :

$$\mathcal{F}_1(u, y_o) = \int_0^u \mathcal{F}_1'(v, y_o)\, dv \qquad (2.17)$$

où $\mathcal{F}_1'(v, y_o)$ est la dérivée supérieure de $\mathcal{F}_1$ au point v (2). Cette dérivée reste évidemment comprise entre 0 et 1. D'autre part, quel que soit ε on peut trouver y_o tel que $\mathcal{F}_1(u, y_o) < \varepsilon^2$ et le théorème de la moyenne nous montre alors que $\mathcal{F}_1'(v, y_o)$ reste inférieur à ε sauf peut-être sur un ensemble de mesure au plus égale à ε. On en déduit aisément que sauf peut-être sur un ensemble de mesure S_2 nulle de valeurs de v

$$\lim_{y \to -\infty} \mathcal{F}_1'(v, y_o) = 0$$

On voit de même que sauf peut-être sur S_2 :

$$\lim_{y \to +\infty} \mathcal{F}_1'(v, y_o) = 1$$

Ceci posé, désignons par $\mathcal{F}_x(y)$ la quantité égale à $\mathcal{F}_1'(v, y_o)$ quand $v = \mathcal{A}(x)$. On a en se reportant à la définition de l'intégrale de Lebesgue :

$$\mathcal{F}(x,y) = \int_{-\infty}^{x} \mathcal{F}_\xi(y)\, d\mathcal{A}(\xi)$$

et on aura en appelant T l'ensemble des points ξ tels que $\xi \leqslant x$:

$$F(x, y) = F_1(x,y) + \mathcal{F}(x, y)$$

$$= \int_{S_1 T} F^*_\xi(y)\, d A_1(\xi) + \int_{(R_1 - S_1 - S_2)\, T} \mathcal{F}_\xi(y)\, d\mathcal{A}(\xi)$$

(1) Fréchet [9] p. 273.

(2) Il est aisé de voir que cette dérivée est une fonction mesurable-B (cf Lavallée Poussin [1] p. 67).

et en posant :

$$(2.18)\ F_x(y) = \begin{cases} F_x^*(y) & \text{si } x = x_i \text{ est un élément de } S_1 \\ \mathcal{F}_x(y) & \text{si } x \text{ est un élément de } (R_2 - S_1 - S_2) \\ H_x(y) & \text{fonction de répartition arbitraire si } x \text{ est un élément de } S_2 \end{cases}$$

on aura bien :

$$F(x,y) = \int F_x(y)\, d\, A(\xi)$$

C.Q.F.D.

En tenant compte de la note de la page précédente, on voit immédiatement qu'on peut choisir $H_x(y)$ de manière que $F_x(y) = \gamma(x,y)$ soit une fonction de la variable x B-mesurable.

D'une façon plus particulière encore si $\delta(x, y, \Delta x)$ défini par 2.12 tend vers une limite unique quand Δx tend vers 0, il est clair que cette limite sera nécessairement égale à $\gamma(x,y)$ si x est un point de l'ensemble $R_2 - S_2$, ce qui prouve que $\left\{\mathcal{G}_x(y)\right\}$ est un système admissible de fonctions de répartition liées $F_x(y)$.

4. — ENSEMBLE ADMISSIBLE DE FONCTIONS CARACTERISTIQUES LIÉES

A tout ensemble admissible E de fonctions de répartition $\left\{F_x(y)\right\}$ nous ferons correspondre un ensemble E' de fonctions caractéristiques liées $\left\{\varphi_x(v)\right\}$ telles que :

$$\varphi_x(v) = \int e^{ivy}\, d\, F_x(y)$$

et nous appellerons E' ensemble admissible des fonctions caractéristiques liées. Comme par définition tout ensemble E de $F_x(y)$ vérifie (2.6) on aura:

$$\int\int e^{i(ux+vy)}\, d\, F(x,y) = \int\int e^{i(ux+vy)}\, d\, F_x(y)\, d\, A(x)$$

ou :

$$\varphi(u,v) = \int e^{iux}\, \varphi_x(v)\, d\, A(x)$$

de même en faisant u = 0, on voit que :

$$\varphi(0,v) = \int_{-\infty}^{+\infty} \varphi_x(v)\, d\, A(x)$$

5. — SURFACES DE PROBABILITÉ REMARQUABLES

A) Deux v.a. X et Y sont dites indépendantes si, quels que soient x et y, on a :

$$F(x,y) = A(x).\ B(y)$$

Nous allons montrer que : Si deux v.a. sont indépendantes, on peut trouver un système admissible de fonctions de répartition liées pour lequel $F_x(y) = B(y)$ quel que soit x et réciproquement.

En effet, si je pose $F_x(y) = B(y)$ j'ai d'après (2.9)

$$F(x,y) = \int B(y)\, d\, A(x)$$
$$= B(y).\ A(x)$$

donc nous avons bien défini un système admissible de fonctions de répartition tel que (2.18) soit vérifiée.

Réciproquement, s'il existe un système admissible de fonctions de répartition tel que $F_x(y) = B(y)$, alors (2.18) sera nécessairement vérifiée.

Il en résulte que la condition nécessaire et suffisante pour que deux variables X et Y soient indépendantes, est qu'il existe un système admissible de fonctions de répartition $\{F_x(y)\}$ telles que :

$$F_x(y) = C(y) \qquad (2.19)$$

B) Ce résultat va nous permettre d'étendre au cas général d'autres notions qui n'ont été jusqu'ici bien définies que si le couple aléatoire X Y admet une densité de probabilité.

Etendant des définitions de Serge Bernstein [1] :

Nous dirons que la v.a. Y est en corrélation dure, relativement à x si on peut trouver un système admissible de fonctions de répartition liées $\{F_x(y)\}$ tel que, quel que soit x, on ait :

$$F_x(y) = C\ [y - \varphi(x)] \qquad (2.20)$$

Nous dirons que la v.a. Y est en corrélation simplement élastique, relativement à x si on peut trouver un système admissible de fonctions de répartition liées $\{F_x(y)\}$ tel que, quel que soit x, on ait :

$$F_x(y) = C\ [\lambda(x).y] \qquad (2.21)$$

plus généralement.

Nous dirons que la v.a. Y est en corrélation isogène relativement à x si on peut trouver un système admissible de fonctions de répartition liées

$$F_x(y) = C\ \{\lambda(x).\ [y - \varphi(x)]\} \qquad (2.22)$$

C) Le concept de fonction de répartition liée va également nous permettre de généraliser la notion de régression linéaire ou parabolique.

Nous dirons par exemple que la courbe de régression des moyennes de Y en x est rectilinéaire s'il existe un système admissible de fonctions de répartition $\{F_x(y)\}$ et deux nombres a et b tels que si nous posons :

$$y(x) = \int_{-\infty}^{+\infty} \eta \quad d\, F_x(\eta) \qquad (2.23)$$

on ait :

$$y(x) = ax + b \qquad (2.24)$$

6. — QUELQUES PROPRIÉTÉS DES SYSTÈMES DE FONCTIONS DE RÉPARTITION ADMISSIBLES, RELATIVEMENT A Ø

1°/ Il est aisé de voir que :

Si la courbe de régression des moyennes de Y en x est rectilinéaire et s'il existe un système de fonctions de répartition $\{\mathcal{F}_x(y)\}$ admissible

relativement à la moyenne $\phi \left\{ F(y) \right\} = \int y \, d F(y)$, alors on a :

$$\phi \left[\mathscr{F}_x (y) \right] = \alpha x + \beta \tag{2.25}$$

en effet, soit $\left\{ F_x(y) \right\}$ un système admissible de fonctions de répartition pour lequel (2.24) est réalisé, on aura :

$$\int_a^b \left[\int_{-\infty}^{+\infty} y \, d F_x(y) \right] d F(x,\infty) = \int_a^b (\alpha x+\beta) \, d F(x,\infty)$$

donc $\phi \left\{ F(y \mid a < x \leqslant b) \right\}$ est compris d'après le théorème de la moyenne entre $\alpha a + \beta$ et $\alpha b + \beta$ et sa limite inférieure $\phi \left[\mathscr{F}_x (y) \right]$ (cf. 2.9) est $\alpha a + \beta$

La démonstration continuera à s'appliquer si on remplace dans 2.24 la droite $\alpha x+\beta$ par une courbe continue quelconque $\varphi(x)$. Cette dernière remarque va nous permettre de généraliser la notion de régression.

Définition :

S'il existe un système (S) de fonctions de répartition $\mathscr{F}_x(y)$ admissible relativement à la fonctionnelle ϕ, nous appellerons courbe associée à la fonctionnelle ϕ (et au système S) la courbe) :

$$\varphi(x) = \phi \left[\mathscr{F}_x (y) \right]$$

En particulier si nous prenons pour $\phi \left\{ F(y) \right\}$ la valeur typique d'ordre k de la v.a. Y de fonction de répartition F(y) alors la courbe $\varphi (x) = \phi \left\{ \mathscr{F}_x (y) \right\}$ sera par définition la ligne de régression de la valeur typique d'ordre k de Fréchet (1)

D'une manière plus particulière encore, si nous prenons $\phi \left[F(y) \right] = \int y \, d F(y)$ la courbe $\varphi (x) = \phi \left[\mathscr{F}_x (y) \right]$ sera la ligne de régression des moyennes

LECTURES RECOMMANDEES

sur les diverses définitions de la probabilité consulter

Blanc Lapierre et Fortet [1] Cramér [1] Feller [1] Paul Lévy [1] [2] Mme Gruzewska [1] [2] Fréchet [9] Lœve [1]

sur les questions de la mesure consulter

Lebesgue [1] La Vallée Poussin [1] Saks [1] Dieudonné [1] [2] [3] Lœve [1] et ensuite seulement Doob [1] [2] Halmos [1]

(1) Il n'est évidemment pas certain que cette ligne de régression soit unique.

CHAPITRE III

NOTION DE GAIN D'INFORMATION

1. — DÉFINITION DE LA NOTION DE GAIN D'INFORMATION

A/ Nous partirons des notions exposées dans le chapitre précédent pour essayer de caractériser l'avantage que nous retirons de la connaissance complète d'une surface de probabilité $F(x,y)$. A cet effet, nous comparerons l'incertitude sur les fonctions de répartition liées $\mathcal{F}_x(y)$ à celle sur la fonction de répartition marginale $F(\infty, y) = B(y)$ et parviendrons à démontrer que :

Théorème 1.

Pour tout système de fonctions de répartition liées $\{\mathcal{F}_x(y)\}$ admissible relativement à la fonctionnelle concave ϕ et tel que $\phi\left[\int \mathcal{F}_x(y)\, d\,\mathcal{A}(x)\right]$ existe quelle que soit la fonction de répartition $\mathcal{A}(x)$, et soit à variation bornée, alors :

$$\phi\left[B(y)\right] \geqslant \int \phi\left[\mathcal{F}_x(y)\right] d\,A(x) \qquad (3.1)$$

Nous poserons :

$$\phi\left[B(y)\right] = \mathcal{I}_Y \qquad \phi\left[\mathcal{F}_x(y)\right] = \mathcal{I}_{Y_x}$$

de sorte que (3.1) s'écrit :

$$\mathcal{I}_Y - \mathcal{M}\mathcal{I}_{Y_x} \geqslant 0 \qquad (3.2)$$

l'inégalité (3.2) montre que si les conditions du théorème I sont vérifiées, l'incertitude relative à la v.a. Y est toujours plus grande ou au plus égale à la moyenne des incertitudes conditionnelles $\mathcal{I}_{Y_x}$. Le fait de connaitre la valeur prise par la v.a. X entrainera donc en moyenne une diminution de l'imprécision de nos connaissances sur la valeur prise par Y. Nous traduirons ce fait fondamental en disant que la connaissance de la valeur prise par X, nous fait perdre de l'incertitude ou ce qui revient au même gagner de l'information - Et nous appellerons gain d'information la quantité

$$\Delta\mathcal{I} = \mathcal{I}_Y - \mathcal{M}\mathcal{I}_{Y_x} = \mathcal{M}\left(-\mathcal{I}_{Y_x}\right) - \left(-\mathcal{I}_Y\right)$$

Lemme auxiliaire I. $F_1(y)$ et $F_2(y)$ étant de la forme $\int \mathcal{F}_x(y)\, d\,\mathcal{A}(x)$ (3.3) la fonctionnelle $\phi\left[\lambda F_1(y) + (1-\lambda) F_2(y)\right]$ existera et sera continue sur le segment de droite $(F_1(y), F_2(y))$.

$\phi\left[\lambda F_1(y) + (1-\lambda) F_2(y)\right]$ existera évidemment car elle est, elle aussi de la forme $\int \mathcal{F}_x(y)\, d\,\mathcal{A}(x)$. De plus $F_1(y)$ et $F_2(y)$ étant fixées

$\phi \left[\lambda F_1(y) + (1-\lambda) F_2(y) \right]$ sera uniquement une fonction $\varphi(\lambda)$ du paramètre λ ($0 \leqslant \lambda \leqslant 1$) on aura :

$$\varphi(\lambda_1) = \phi \left[\lambda_1 F_1(y) + (1-\lambda_1) F_2(y)\right]$$
$$\varphi(\lambda_2) = \phi \left[\lambda_2 F_1(y) + (1-\lambda_2) F_2(y)\right]$$

Ecrivons que ϕ est une fonctionnelle concave, il vient :

$$\frac{1}{2} \left\{ \phi \left[\lambda_1 F_1(y) + (1-\lambda_1) F_2(y)\right] + \phi\left[\lambda_2 F_1(y) + (1-\lambda_2) F_2(y)\right] \right. \leqslant \phi\left[\frac{\lambda_1+\lambda_2}{2} F_1(y) + \left(1-\frac{\lambda_1+\lambda_2}{2}\right) F_2(y)\right]$$

ou :

$$\frac{1}{2}\left[\varphi(\lambda_1) + \varphi(\lambda_2)\right] \leqslant \varphi\left[\frac{\lambda_1+\lambda_2}{2}\right]$$

donc $\varphi(\lambda)$ est elle aussi une fonction concave de λ ; elle est de plus évidemment <u>bornée</u>, donc elle est <u>continue</u> (1)

Plus généralement $\phi \left[\lambda_1 F_1(y) + \lambda_2 F_2(y) + \cdots + \lambda_p F_p(y)\right]$ où $\lambda_1 + \lambda_2 + \ldots + \lambda_p = 1$ est continue sur la portion de l'hyperplan $\left[F_1(y), F_2(y), \ldots F_p(y)\right]$ telle que :

$$\lambda_1 + \lambda_2 + \ldots + \lambda_p = 1 \qquad\qquad 0 \leqslant \lambda_i \leqslant 1$$

on le voit aisément en posant $f(y) = \dfrac{1}{1-\lambda_p} \displaystyle\sum_1^{p-1} \lambda_i F_i(y)$

<u>Lemme II. Quels que soient le nombre</u> $p_1, p_2, \ldots p_m$ $0 \leqslant p_i \leqslant 1$ $\sum p_i = 1$ <u>et les fonctions</u> $F_i(y)$ <u>de la forme (3.3), on a</u> :

$$\sum_{i=1}^{m} p_i\, \phi\left[F_i(y)\right] \leqslant \phi\left[\sum_1^m p_i F_i(y)\right] \tag{3.4}$$

<u>Démonstration</u>

Par définition, on a : $\frac{1}{2}\left\{\phi\left[F_1(y)\right] + \phi\left[F_2(y)\right]\right\} \leqslant \phi\left[\frac{F_1(y)+F_2(y)}{2}\right]$

Dès lors si $F_1(y)$, $F_2(y)$, $F_3(y)$, $F_4(y)$ sont 4 fonctions de répartition de la forme (3.3), on a :

$$\frac{1}{4}\left\{\phi\left[F_1(y)\right] + \phi\left[F_2(y)\right] + \phi\left[F_3(y)\right] + \phi\left[F_4(y)\right]\right\}$$

$$\leqslant \frac{1}{2}\phi\left[\frac{F_1(y)+F_2(y)}{2}\right] + \frac{1}{2}\phi\left[\frac{F_3(y)+F_4(y)}{2}\right]$$

$$\leqslant \phi\left[\frac{F_1(y)+F_2(y)+F_3(y)+F_4(y)}{4}\right]$$

plus généralement :

$$2^{-r}\sum_{\nu=1}^{2^r}\phi\left[F_\nu(y)\right] \leqslant \phi\left[2^{-r}\sum_{\nu=1}^{2^r} F_\nu(y)\right]$$

(1) Hardy Littellewood et Polya [1] p. 93.

En particulier, si $F_1(y)$, $F_2(y)$ ... $F_n(y)$ sont données et si nous faisons dans la formule précédente :

$$F_{n+1}(y) = F_{n+2}(y) = \qquad = F_{2^r}(y) = \frac{F_1(y) + F_2(y) + \ldots + F_n(y)}{n}$$

il vient :

$$2^{-r}\left\{\sum_1^n \phi\left[F_\nu(y)\right] + (2^r - n)\,\phi\left[\frac{1}{n}\sum_1^n F_\nu(y)\right]\right\} \leqslant \phi\left[\frac{1}{n}\sum_1^n F_\nu(y)\right]$$

ou :

$$\frac{1}{n}\sum_1^n \phi\left[F_\nu(y)\right] \leqslant \phi\left[\frac{1}{n}\sum_1^n F_\nu(y)\right] \qquad (3.5)$$

Dans le cas particulier où n_1 des $F_\nu(y)$ sont identiques à $F_1(y)$, n_2 identiques à $F_2(y)$ et n_m identiques à $F_m(y)$ $(n_1 + n_2 + n_m = n)$ (3.5) prend la forme :

$$\frac{1}{n}\left\{n_1\,\phi\left[F_1(y)\right] + n_2\,\phi\left[F_2(y)\right] + \ldots + n_m\,\phi\left[F_m(y)\right]\right.$$

$$\leqslant \phi\left[\frac{n_1}{n}F_1(y) + \ldots + \frac{n_m}{n}F_m(y)\right] \qquad (3.6)$$

Cette dernière inégalité démontre que, quels que soient les nombres p_i rationnels, on a :

$$\sum_1^m p_i\,\phi\left[F_i(y)\right] \leqslant \phi\left[\sum_1^m p_i\,F_i(y)\right] \qquad (3.6)$$

Or, de la généralisation du théorème I résulte que ϕ est continue, donc (3.6) est vraie pour toute valeur de p_i rationnelle ou non, ce qui démontre le lemme II.

Pour achever la démonstration du théorème I, remarquons que $\phi\left[\mathcal{F}_x(y)\right]$ est certainement mesurable-B.

En effet, considérons la fonction D(e) égale à $\phi\left[\mathcal{F}(y \mid e)\right]$ si cette quantité est définie et égale à m dans le cas contraire, et attachons à chaque grille G_n un ensemble de nombres $D(e_n)$ et posons $D_n = D(e_n)$ cette fonction D_n est mesurable-B, car elle est constante dans chaque maille ω_n et $\phi\left[\mathcal{F}_x(y)\right]$ qui est la plus petite limite de D est aussi mesurable-B.

De plus, d'après le lemme II, on a, en choisissant convenablement les $F_i(y)$:

$$\sum_i \phi\left[F(y \mid e_i)\right] \Pr(e_i) \leqslant B(y) \qquad (3.7)$$

où les e_i sont disjoints et tels que $\sum \Pr(e_i) = 1$

Il suffira pour achever de prouver le théorème I, d'appliquer le théorème de Fatou. On a en effet :

$$\int \text{lim. inf. } D_n(x)\, d\,F(x, \infty) \geqslant \int \underset{n}{\text{lim. inf.}}\, D_n(x) . d\,F(x, \infty)$$

$$\geqslant \int \phi\left[\mathcal{F}_x(y)\right] . d\,F(x, \infty)$$

inégalité qui prouve (3.1).

B/ Généralisation de la notion de gain d'information

La notion de gain d'information précédemment définie par (3.2) présente l'inconvénient de n'être définie que s'il existe un système de fonctions de répartition liées admissibles relativement à la fonctionnelle ϕ.

Or, si nous sommes bien arrivés à prouver qu'il existe toujours un système admissible de fonctions de répartition liées, en revanche, on peut construire des exemples où $\phi\left[F(y) \mid e_i\right)]$ existe bien, quel que soit e_i mais où néanmoins on ne peut construire de système admissible, relativement à la fonctionnelle ϕ.

Aussi généraliserons-nous la notion de gain d'information en donnant une définition de cette quantité qui ne fait intervenir que la notion de probabilité liée à un intervalle (sans faire intervenir la notion de système admissible lié à un point).

Nous aurons besoin dans nos démonstrations de considérer les

$$F_k(y) = \sum \lambda_i F(y \mid e_i) \tag{3.8}$$

où

$$\sum \lambda_i = 1$$

et nous énoncerons le théorème suivant :

Théorème II.

Si quelque soit $F_k(y)$ défini par (3.8), la fonctionnelle concave $\phi\left[F_k(y)\right]$ existe et reste comprise entre deux nombres donnés m et M, alors, quels que soient les intervalles disjoints e_i tels que $\sum \Pr(e_i) = 1$, on a :

$$\phi\left[B(y)\right] - \sum \phi\left[F(y \mid e_i)\right] \Pr(e_i) \geqslant 0 \tag{3.9}$$

La démonstration de ce théorème est absolument identique à celle du lemme II. Il suffit pour s'en convaincre de lire la démonstration précédente en donnant aux $F_k(y)$ leur nouveau sens.

Dès lors, nous appellerons gain d'information la borne supérieure du premier membre de 3.9 et poserons :

$$\Delta \mathfrak{I} = \text{borne sup.} \left\{ \phi\left[B(y)\right] - \sum \phi\left[F(y \mid e_i)\right] \Pr(e_i) \right\} \tag{3.9}$$

Remarque.

La condition de concavité, qui figure dans le théorème II, n'est pas une condition accessoire. En effet, la condition nécessaire et suffisante pour que la quantité :

$$\text{borne sup.} \left\{ \phi\left[B(y)\right] - \sum_i \phi\left[F(y \mid e_i)\right] \Pr(e_i) \right\} \tag{3.10}$$

soit non-négative, pour tous les couples aléatoires X Y pour lesquels elle est définie, est que ϕ soit une fonctionnelle concave.

En effet, nous venons de voir que la condition est suffisante. Pour montrer qu'elle est nécessaire, supposons qu'il existe deux fonctions de répartition $F_1(y)$ et $F_2(y)$ telles que :

$$\frac{1}{2}\left\{ \phi\left[F_1(y)\right] + \phi\left[F_2(y)\right] \right\} > \phi\left[\frac{F_1(y) + F_2(y)}{2}\right]$$

alors si nous considérons le couple aléatoire X Y tel que :

$$\Pr(X = x_1,\ Y \leqslant y) = \frac{1}{2} F_1(y)$$

$$\Pr(X = x_2,\ Y \leqslant y) = \frac{1}{2} F_2(y)$$

$$\Pr (X \text{ différent de } x_1 \text{ et de } x_2) = 0$$

il est clair que pour ce couple la quantité (3.10) sera bien définie et vaudra :

$$\phi \left[\frac{F_1(y) + F_2(y)}{2}\right] - \frac{1}{2} \left\{ \phi \left[F_1(y) \right] + \phi \left[F_2(y) \right] \right\}$$

quantité qui est négative.

2. — CAS DE L'INDÉPENDANCE

Nous avons vu au chapitre précédent que si 2 variables X et Y sont indépendantes, alors $F(y \mid e_i) = B(y)$ toutes les fois que $\Pr(e_i) \neq 0$, dès lors on voit immédiatement que la quantité :

$$\phi \left[B(y)\right] - \sum \phi \left[F(y \mid e_i)\right] . \Pr(e_i)$$

est nulle quels que soient les e_i, sa limite $\Delta \mathfrak{I}$ sera donc nulle, donc si X et Y sont indépendantes, le gain d'information est nul

A/ La réciproque n'est pas vraie. Il peut en effet arriver que X et Y ne soient pas indépendantes, bien que $\Delta \mathfrak{I}$ soit nulle. Il en sera par exemple ainsi lorsque quelque soit e_i, $\phi \left[B(y)\right] = \phi \left[F(y \mid e_i)\right]$

1) Exemple I.

Considérons le cas où la variance est prise comme mesure de l'incertitude et supposons que quel que soit e la fonction de répartition $F(y \mid e)$ ait une moyenne m et une variance σ^2 donnés d'avance ; dans ce cas on aura bien :

$$\phi \left[B(y)\right] = \phi \left[F(y \mid e)\right]$$

Cette propriété qui résulte immédiatement de l'hypothèse faite, résulte aussi du fait que si on considère un ensemble d'intervalles e disjoints, tels que :

$$\sum \Pr (e_j) = 1 \tag{3.11}$$

on aura en notant $\sigma^2_{e_j}$ et m_{e_j} la variance et la moyenne de la v.a. de fonction et répartition $F(y \mid e_j)$

$$\overline{Y} = \int y \, d\, B(y) = \sum m_{e_j} \Pr(e_j) = m_{e_j}$$

$$\sigma^2_Y = \int (y - \overline{Y})^2 \, d\, B(y) = \sum \left[\sigma^2_{e_j} + (m_{e_j} - \overline{Y})^2\right] \Pr(e_j) \tag{3.12}$$

Dès lors si les fonctions de répartition $F(y \mid e_j)$ ne sont pas identiques, bien qu'ayant même moyenne et même variance X et Y ne seront pas indépendantes. Et pourtant on a bien $\Delta \mathfrak{I} = 0$, si par exemple on suppose en outre que $F(y \mid e_i) = F(y \mid e_j)$ pour tout $e_i \subset e_j$.

2) Généralisations

a) L'exemple précédent est instructif car il nous montre que si nous supposons que $F(y \mid e)$ a une moyenne constante alors quels que soient les e_j vérifiant (3.11), (3.12) nous montre que :

$$\sigma^2_Y = \sum \sigma^2_{e_j} \Pr (e_j) \tag{3.13}$$

et en passant à la limite on voit que $\Delta \mathfrak{I} = 0$

Réciproquement, si sur un ensemble e_j tel que $\Pr e_j \neq 0$ on a $m_{e_j} \neq \overline{Y}$ alors (3.12) nous montre qu'on aura :

$$\Delta \mathfrak{I} \geqslant \Sigma (m_{e_j} - \overline{Y})^2 \Pr (e_j)$$

La condition nécessaire et suffisante pour que $\Delta \mathfrak{I} = 0$ quand on prend la variance pour mesure de l'incertitude, est donc que $\int y \, dF(y \mid e)$ soit indépendante de e pour tout intervalle e pour lequel $F(y \mid e)$ est défini.

b) Plus généralement si nous prenons le moment typique d'ordre k de Fréchet $m^{(k)}$ comme mesure de l'incertitude, on aura en posant :

$$m_Y^{(k)} = \min_a . \int \left| y-a \right|^k dB(y) \tag{3.14}$$

$$m_{(e)}^{(k)} = \min_b . \int \left| (y-b \right|^k dF(y \mid e) \tag{3.15}$$

et en supposant que les e_j vérifiant (3.11) :

$$m_Y^{(k)} = \min_a . \sum_j \left\{ \int \left| y-a \right|^k dF(y \mid e_j) \right\} \Pr(e_j)$$

et l'on aura toujours évidemment :

$$m_Y^{(k)} \geqslant \sum_j \left\{ \min_b . \int \left| y-b \right|^k dF(y \mid e_j) \right\} \Pr(e_j) \tag{3.16}$$

on ne pourra évidemment avoir le signe = dans (3.16) que si le minimum de chaque intégrale $\int \left| y-b_j \right|^k dF(y \mid e_j)$ est atteint pour la valeur a pour chaque e tel que $\Pr(e_j) \neq 0$. Autrement dit, si a est valeur typique d'ordre k de la v.a. Y_j qui a pour fonction de répartition $F(y \mid e_j)$.

En raisonnant comme précédemment, on arrive ainsi à montrer que :

La condition nécessaire et suffisante pour que $\Delta \mathfrak{I} = 0$ quand on prend le moment typique d'ordre k de Fréchet comme mesure de l'incertitude, est qu'il existe une constante a qui soit valeur typique d'ordre k de toutes les v.a. qui ont une fonction de répartition de la forme $F(y \mid e)$ (toutes les fois que $F(y \mid e)$ est défini).

En particulier :

La condition nécessaire et suffisante pour que $\Delta \mathfrak{I} = 0$ quand on prend l'écart moyen $m^{(1)} = \min_\alpha . \int \left| y-\alpha \right| dF(y)$ comme mesure de l'incertitude est qu'il existe une valeur a qui soit égale à l'une des médianes des v.a. Y_e de fonction de répartition $F(y \mid e)$.

B/ Fonctionnelle strictement concave

1/ Définition :

Nous dirons encore que ϕ est une fonctionnelle strictement concave si

$$\phi \left[\frac{F_1(y) + F_2(y)}{2} \right] > \frac{1}{2} \left\{ \phi \left[F_1(y) \right] + \phi \left[F_2(y) \right] \right\} \tag{3.17}$$

pour tout couple de fonctions de répartition $F_1(y)$ et $F_2(y)$ non identiques.

Autrement dit, on ne peut avoir le signe égal dans l'inégalité (1.3) que si $F(y) = G(y)$.

Dans ce cas, on voit en reprenant les calculs du début de ce chapitre que l'on ne peut avoir le signe égal dans (3.9) que si tous les $F(y \mid e_i)$ sont identiques. On en déduit le théorème suivant :

Théorème :

Si on prend une fonctionnelle strictement concave comme mesure de l'incertitude, on aura $\Delta \mathcal{J} = 0$ si et seulement si X et Y sont indépendantes.

Au contraire, si ϕ est une fonctionnelle concave sans être une fonctionnelle strictement concave, il existe des couples aléatoires X Y tels que $\Delta \mathcal{J} = 0$, sans que X et Y soient indépendantes.

2/ Application à l'entropie

Supposons que le couple X Y soit susceptible de ne prendre qu'un nombre fini de valeurs (x_i, y_i) et soit p_{ij} la probabilité du couple x_i, y_j. Dès lors, si nous prenons pour ϕ l'entropie, on aura :

$$\phi \left[F(y) \right] = - \sum p_j \, L \, p_j \tag{3.18}$$

Cette fonction étant définie, quelles que soient les probabilités p_j des y_j telles que $\sum p_j = 1$, il est clair que $\phi \left[\sum \lambda_i F_{x_i}(y) \right]$ où $\sum \lambda_i = 1$, sera bien définie.

Or, ϕ définie par (3.18), est une fonctionnelle strictement concave. En effet si nous considérons une autre f.d.r. $F^{*}(y)$ pour laquelle Y prendra les mêmes valeurs y_j avec des probabilités p'_j ($\sum p'_j = 1$) on aura :

$$\phi \left[\frac{F(y) + F^{*}(y)}{2} \right] = - \sum_j \frac{p_j + p'_j}{2} \, L \, \frac{p_j + p'_j}{2}$$

or $- x \, Lx$ étant une fonction strictement concave, on aura :

$$- \frac{p_j + p'_j}{2} \, L \, \frac{p_j + p'_j}{2} \geqslant - \frac{1}{2} \left[p_j \, L \, p_j + p'_j \, L \, p'_j \right] \tag{3.19}$$

et l'on ne pourra avoir le signe égal dans (3.19) que si $p_j = p'_j$

Il en résulte qu'on aura toujours :

$$- \sum_j \frac{p_j + p'_j}{2} \, L \, \frac{p_j + p'_j}{2} \geqslant - \frac{1}{2} \left\{ \sum_j p_j \, L \, p_j + \sum_j p'_j \, L p'_j \right\} \tag{3.20}$$

et l'on ne pourra avoir le signe égal dans (3.20) que si tous les p_j sont égaux aux p'_j.

Autrement dit, on ne pourra avoir :

$$\phi \left[\frac{F(y) + F^{*}(y)}{2} \right] = \frac{1}{2} \left\{ \phi \left[F(y) \right] + \phi \left[F^{*}(y) \right] \right\}$$

que si $F(y) \equiv F^{*}(y)$

Donc, si X et Y sont des v.a. discontinues susceptibles de prendre un nombre fini de valeurs, et si on prend l'entropie comme mesure de l'incertitude, on aura $\Delta \mathcal{J} = 0$ si et seulement si X et Y sont indépendantes.

b) Considérons maintenant le cas où on prend l'entropie comme mesure de l'incertitude sur la v.a. Y possédant une densité de probabilité $f(y)$ $f(y) = F'(y))$ bien définie et bornée en tout point, et posons :

$$\phi \left[F(y) \right] = - \int f(y) \, L \, f(y) \, dy \tag{3.21}$$

et supposons que pour un couple aléatoire XY, ϕ soit définie pour toutes les fonctions $F_k(y)$ définies par (3.8). (Il en sera notamment ainsi si $\frac{\partial^2 F(x,y)}{\partial x \, \partial y} = f(x,y)$ existe et est continue en tout point du plan et si de plus, pour tout x tel que :

$$f_x^*(y) = \frac{f(x,y)}{\int f(x,\eta)\, d\eta} \quad \text{existe}$$

l'intégrale :

$$\int f_x^*(y)\, L\, f_x^*(y)\, dy \tag{3.22}$$

existe également).

Dans ces conditions, si $F_1(y)$ et $F_2(y)$ sont définis par (3.8) $\frac{F_1(y) + F_2(y)}{2}$ le sera aussi, donc $\phi\left[\frac{F_1(y) + F_2(y)}{2}\right]$ existera.

De plus, on aura évidemment si on pose :

$$F'_1(y) = f_1(y) \qquad F'_2(y) = f_2(y)$$

$$-\frac{f_1(y) + f_2(y)}{2}\, L\, \frac{f_1(y) + f_2(y)}{2} \geqslant -\frac{1}{2}\left[f_1(y) L f_1(y) + f_2(y) L f_2(y)\right]$$

et l'on ne pourra avoir le signe égal que si $f_1(y) = f_2(y)$.

Dès lors, on voit aisément que l'on ne pourra avoir :

$$\phi\left[\frac{F_1(y) + F_2(y)}{2}\right] = \frac{1}{2}\left\{\phi\left[F_1(y)\right] + \phi\left[F_2(y)\right]\right\} \tag{3.23}$$

que si $f_1(y)$ et $f_2(y)$ sont identiques, sauf peut-être sur un ensemble de mesure nulle E_1.

On aura donc puisque E_1 est de probabilité nulle :

$$F_1(y) = F_2(y)$$

Donc dans ce cas, on aura encore $\Delta \mathfrak{I} = 0$ si et seulement si X et Y sont indépendants.

3. — NOTION DE FONCTION UNIVALENTE

La définition la plus naturelle qui vient à l'esprit est de dire que Y est fonction univalente de x si la valeur de Y est un nombre certain $\varphi(x)$ quand on connait x. Mais de même qu'on est amené en calcul des probabilités à substituer à la notion analytique de convergence, d'une suite vers un nombre a, les notions de convergence en probabilité, de convergence en moyenne d'ordre k, de convergence presque sûre, de même nous serons aussi amenés à élargir la notion de fonction univalente de l'analyse.

Diverses définitions sont alors possibles :

Nous dirons que Y est une fonction univalente de x du type A. Si on peut trouver un système admissible de fonctions de répartition liées $\{F_x(y)\}$ pour lequel quel que soit x il existe une valeur $y_0(x)$ telle que :

$$F_x(y) = \begin{cases} 0 \text{ pour } y < y_0(x) \\ 1 \text{ pour } y \geqslant y_0(x) \end{cases}$$

Mais, en réalité, nous avons vu que nous obtenions des résultats plus généraux en nous passant de la notion de système admissible de fonction de répartition liée.

A cet effet, considérons un réseau linéaire sur l'axe 0x défini par un nombre illimité de grilles G_1, G_2... G_n. La grille G_2 sera obtenue à partir de la grille G_1 en divisant chaque maille de la grille G_1 en k partie égales. Les mailles correspondant à la grille G_n seront des intervalles semi-ouverts à droite et nous noterons par ℓ_n une maille quelconque de G_n. Ainsi, un point x appartiendra à une maille ℓ_n° de G_n et une seule de ces mailles. $\ell_1^\circ, \ell_2^\circ \ldots \ell_n^\circ \ldots$ sont emboitées les unes dans les autres.

Ceci posé, nous dirons de Y est fonction univalente de x au point x_0 en probabilité (en moyenne d'ordre k, presque sûrement

Si 1°/ Quel que soit n $F(y|\ell_n^\circ)$ est bien défini

2°/ si Y_{ℓ° désignant la v.a. de fonction de répartition $F(y|\ell_n^\circ)$

$Y_{\ell_n^\circ}$ converge en probabilité (en moyenne d'ordre k presque sûrement) vers une constante (1).

Il est bien clair avec cette définition que si Y est fonction univalente de x au point x_0 au sens de l'analyse elle l'est à l'un quelconque des sens stochastiques ci-dessus. De plus, il est bien évident que si Y est fonction univalente de x au point x en moyenne quadratique ou presque sûrement elle l'est aussi en probabilité.

De même, nous dirons que Y est fonction univalente de x en probabilité (en moyenne d'ordre k, presque sûrement) si c'est une fonction univalente de x pour un ensemble E de points x_0 dont la probabilité totale est unité.

Nous dirons de plus que Y est fonction univalente de x au sens strict : a) en probabilité, b) en moyenne d'ordre k, c) presque sûrement. Si cette convergence est en quelque sorte uniforme, c'est-à-dire si à tout couple de nombres positifs ε et η on peut faire correspondre un nombre N tel que pour tout $n > N$ et tout x_0 on ait :

$$\text{a) } \max_a \int_{a-\varepsilon}^{a+\varepsilon} dF(y|\ell_n^\circ) > 1-\eta$$

$$\text{b) } \min_a \int_{-\infty}^{+\infty} |y-a|^k \, dF(y|\ell_n^\circ) < \varepsilon$$

c) on peut trouver a tel que les inégalités :

$$\int_{a-\varepsilon}^{a+\varepsilon} dF(y|\ell_\nu^\circ) > 1-\eta \qquad \nu = n, \ n+1, \ldots$$

soient simultanément vérifiées.

Ceci posé, prenons le moment typique d'ordre k de Fréchet comme mesure de l'incertitude. Il est bien évident que si Y est fonction univalente de x au sens strict en moyenne d'ordre k, alors, quel que soit Σ on peut trouver un nombre N tel que $n > N$ entraine :

$$\phi\left[F(y|\ell_n)\right] = \text{borne inf.}_a \int |y-a|^k \, dF(y|\ell_n) < \varepsilon$$

(1) La définition que nous proposons présente toutefois un inconvénient. Il n'est pas prouvé que la définition de la fonction univalente stochastique soit indépendante du système de grilles choisi.

et ceci sera vrai pour tout intervalle ℓ_n tel que $n > N$ et que $P(X \subset \ell_n) \neq 0$ dès lors on aura évidemment :

$$\sum \phi\left[F(y \mid \ell_n)\right] P(\ell_n) < \varepsilon$$

la sommation étant étendue à tous les intervalles de la grille G_n et par conséquent :

$$\text{borne inf.} \sum \phi\left[F(y \mid \ell_i)\right] P(\ell_i) = 0$$

d'où on tire d'après (3.9) :

$$\Delta \mathfrak{I} = \phi\left[B(y)\right]$$

par conséquent si Y est fonction univalente de x en moyenne d'ordre k au sens strict et si on prend l'écart typique d'ordre k de Fréchet comme mesure de l'incertitude, on a :

$$\Delta \mathfrak{I} = \phi\left[B(y)\right]$$

En particulier, si Y est fonction univalente de x en moyenne quadratique au sens strict et si on prend la variance comme mesure de l'incertitude, on a :

$$\Delta \mathfrak{I} = \sigma_Y^2$$

Réciproquement, supposons qu'on prenne le moment typique d'ordre k de Fréchet comme mesure de l'incertitude et qu'on ait :

$$\Delta \mathfrak{I} = \phi\left[B(y)\right]$$

d'après (3.9), on a :

$$\text{borne inf.} \sum \phi\left[F(y \mid e_i)\right] Pr(e_i) = 0$$

Il est alors clair que quel que soit ε on peut trouver des ensembles e_i tels que $\sum \phi\left[F(y \mid e_i)\right] Pr(e_i) < \varepsilon$ et on pourra supposer puisque $\phi\left\{F(y \mid \ell]\right\}$ est supposé borné que les extrémités des e_i ne sont pas points de discontinuité de $F(x, \infty)$.

Considérons les intervalles semi-ouverts α_n^{ij} formés par les noeuds de la grille G_n et par les extrémités des e_i, il est clair d'après le théorème II du présent chapitre que l'on a :

$$\sum \phi\left[F(y \mid \alpha_n^{ij})\right] Pr(X \subset \alpha_n^{ij}) < \varepsilon$$

et sauf peut-être pour un ensemble de α_n^{ij} de probabilité totale inférieure à $\sqrt{\varepsilon}$ on aura $\phi\left[F(y \mid \alpha_n^{ij})\right] < \sqrt{\varepsilon}$ c'est-à-dire que sauf peut-être pour un ensemble de valeurs de x de probabilité nulle, on aura pour tout point x_0

$$\lim_{n\to\infty} \min_{a} \int |y-a|^k \, dF(y \mid \alpha_n^o) = 0$$

α_n^o désignant celui des α_n^{ij} qui contient x_0 autrement dit si $\Delta \mathfrak{I} = \phi\left[B(y)\right]$ quand on prend la valeur typique d'ordre k de Fréchet comme mesure de l'incertitude, Y est fonction univalente de x en moyenne d'ordre k.

En particulier, si $\Delta \mathfrak{I} = \sigma_Y^2$ quand on prend la variance comme mesure de l'incertitude Y est fonction univalente de x en moyenne quadratique.

LECTURES RECOMMANDEES

Le terme gain d'information apparait dans WIENER [1] sur les questions de convexité consulter HARDY LITTLEWODD et POLYA [1] .

CHAPITRE IV

SUR UN SCHÉMA PROBABILISTE

1. — DESCRIPTION D'UN COUPLE ALÉATOIRE QUELCONQUE X Y

Les surfaces de probabilités que l'on rencontre dans la pratique semblent, à beaucoup de statiticiens, trop compliquées pour être étudiées directement. Aussi essaient-ils de les remplacer par des surfaces de probabilité plus simples.

A/ CORRELATION DURE.

Dans le cas particulier où le couple aléatoire XY admet une densité de probabilité $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \ \partial y}$, S. Bernstein [1] a proposé un modèle simple qui englobe la plupart des modèles considérés avant lui. Il écrit :

$$f(x,y) = a(x) \cdot \alpha_x(y) \qquad (4.1)$$

et dit que X et Y sont en corrélation dure si :

$$\alpha_x(y) = \Psi \left[y + \varphi(x) \right]$$

Autrement dit, les courbes $\alpha_x(y)$ sont superposables par une translation.

Géométriquement parlant, une densité de probabilité peut être considérée comme une certaine masse répartie sur le plan des xy avec la densité $f(x,y)$. $\alpha_x(y)$ représente alors la densité sur les fibres élémentaires parallèles à Oy et le fait qu'il y a corrélation dure se traduit en disant que les fibres élémentaires se déduisent les unes des autres par une translation.

Nous dirons que la densité $f(x,y)$ est alors produite par des fibres élémentaires réparties avec la densité $a(x)$ sur l'axe des x et, engendrées par la translation de la distribution de masse Ψ (y) le long de la courbe φ (x).

Nous avons, au chapitre II, généralisé cette définition et avons dit que X et Y sont en corrélation dure si l'on peut trouver un système admissible de fonctions de répartition liées tel que :

$$F_x(y) = \Psi \left[y + \varphi(x) \right] \qquad (4.2)$$

On voit aisément qu'alors on a :

$$\Pr \left[Y \leq \varphi(x) + h, \ X \subset T \right] = \Pr (X \subset T). \Psi \ (h) \qquad (4.3)$$

et réciproquement si (4.3) est vérifiée, quel que soit E, $F_X(y) = \Psi\left[y + \varphi(x)\right]$ est un système admissible de fonctions de répartition.

X et Y seront donc en corrélation dure si et seulement si (4.3) est vérifiée.

B/ SCHEMA PROBABILISTE.

Etant donné un couple aléatoire XY (supposé défini par P(E)) et une fonction $y = \omega(x)$ nous pouvons essayer de remplacer le couple aléatoire XY par un autre couple aléatoire $X^* Y^*$ dans lequel $X^* Y^*$ sont en corrélation dure.

Le couple aléatoire $X^* Y^*$ sera donc supposé défini par $P^*(E)$ supposée telle que :

$$P^*\left[Y \leqslant \omega(x) + h,\ X \subset T\right] = P(X \subset T)\,\Psi(h) \qquad (4.4)$$

$$P^*\left[Y \leqslant \omega(x) + h,\right] = P\left[Y \leqslant \omega(x) + h\right] \qquad (4.5)$$

et ceci quels que soient T et h.

En faisant $T = R_1$ on voit que $\Psi(h)$ est entièrement déterminé par (4.5) puisque (4.4) donne alors :

$$\Psi(h) = P^*\left[Y \leqslant \omega(x) + h\right] \qquad (4.6)$$

2. — GAIN D'INFORMATION DURE ET ÉLASTIQUE

A/ DEFINITION

Nous pouvons caractériser la valeur du schéma probabiliste au moyen de l'incertitude.

Nous avons vu en effet cf. (3.9) que l'on a $\Delta \mathfrak{I} \geqslant 0$

avec : $$\Delta \mathfrak{I} = \phi\left[B(y)\right] - \text{borne inf.} \sum_i \phi\left[F(y \mid e_i)\right] Pr(e_i) \qquad (4.7)$$

et ceci quelle que soit la fonctionnelle concave ϕ.

On peut écrire (4.7) :

$$\Delta \mathfrak{I} = \phi\left[B(y)\right] - \phi\left[\Psi(y)\right] + \phi\left[\Psi(y)\right] - \text{borne inf.} \sum \phi\left[F(y \mid e_i)\right] Pr(e_i) \qquad (4.8)$$

$$= \quad D \quad + \quad E \qquad (4.9)$$

En posant :

$$F(y \mid e) = \frac{P(Y \leqslant y,\ X \subset e)}{P(X \subset e)} \qquad (4.10)$$

$$D = \phi\left[B(y)\right] - \phi\left[\Psi(y)\right]$$

$$E = \phi\left[\Psi(y)\right] - \text{borne inf.} \sum \phi\left[F(y \mid e_i)\right] Pr(e_i)$$

Nous appellerons $\Delta \mathfrak{I}$ gain d'information totale, D gain d'information dure et E gain d'information élastique, et dirons que le gain d'information totale est la somme du gain d'information dure et du gain d'information élastique.

On voit donc que ces quantités sont définies quelle que soit la fonctionnelle concave ϕ

B/ PROPRIETES DU GAIN D'INFORMATION ELASTIQUE ET DU GAIN D'INFORMATION DURE.

1) Gain d'information élastique.

On a, compte tenu de (4.4) et (4.5) :

$$\Psi(y) = P\left[Y \leqslant \omega(x) + y\right] \tag{4.11}$$

donc $\Psi(y)$ est bien une fonction de répartition.

Posons :

$$F_1(y \mid e) = \frac{P\left[Y \leqslant \omega(x) + y,\ X \subset e\right]}{P(X \subset e)} \quad \text{si } P(X \subset e) \neq 0$$

$F_1(y \mid e) = \mathcal{F}(y)$ dans le cas contraire (où $\mathcal{F}(y)$ est une fonction de répartition arbitraire mais telle que $\phi\left[\mathcal{F}(y)\right]$ est défini et borné. $F_1(y \mid e)$ est évidemment une fonction de répartition et on a pour les e_i disjoints et tels que $\sum e_i = R_1$

$$\sum F_1(y \mid e_i)\, P(e_i) = \Psi(y)$$

On a d'après (3.6)

$$\phi\left[\Psi(y)\right] \geqslant \sum \phi\left[F_1(y \mid e_i)\right] P(e_i) \tag{4.12}$$

Il en résulte que :

si quel que soit ε on peut trouver des nombres e_i tels que $e_j \subset e_l$ entraine

$$\phi\left[F(y \mid e_i)\right] - \phi\left[F_1(y \mid e_i\right] < \varepsilon \tag{4.13}$$

alors le gain d'information élastique E ne peut être négatif.

La condition (4.13) est absolument essentielle et nous supposerons toujours dans la suite que nous nous bornons à l'étude des fonctionnelles ϕ et des courbes $\omega(x)$ pour lesquelles (4.13) est vérifiée.

On remarquera que dans le cas où X est une v.a. discontinue pouvant prendre les valeurs $x_1, x_2 \ldots x_n$ (c'est le seul cas où la fonction de répartition liée est définie au sens classique pour des ensembles e_i tels que $\sum Pr(e_i) = 1$), (4.13) est une conséquence de (1.2) - (4.13) s'écrit en effet alors :

$$\phi\left\{F_{x_i}\left[y + \omega(x_i)\right]\right. = \phi\left[F_{x_i}(y)\right] \tag{4.14}$$

2) Gain d'information dure.

Le gain d'information dure, au contraire, peut être positif ou négatif. Le fait que D est négatif signifie que le schéma probabiliste considéré est très mal adapté au problème considéré. D mesure en quelque sorte le gain d'information apporté sur le couple aléatoire X Y par le schéma probabiliste envisagé. Il est d'ailleurs clair que nous sommes tout à fait bien informés sur X Y quand on connait toutes les fonctions de répartition liées (cas auquel le gain d'information est $\Delta\mathfrak{J}$) au contraire la donnée du couple aléatoire X^* Y^* ne nous renseigne que partiellement.

Et en effet, si (4.13) est vérifiée, on a toujours en vertu de (4.9) :

$$D \leqslant \Delta\mathfrak{J} \tag{4.15}$$

3. — CAS OU L'INFORMATION ÉLASTIQUE EST NULLE

A/ SI X ET Y SONT EN CORRELATION DURE ET SI 4.13 EST VERIFIEE, ON A E = 0, si on prend pour $\omega(x)$ la ligne de régression $\varphi(x)$.

En effet (4.3) est alors vérifiée.

On aura donc pour tout ensemble T tel que $Pr(T) \neq 0$

$$\Psi(y) = F_1(y \mid T)$$

donc quels que soient les e_i tels que $\sum e_i = R_1$

$$\phi\left[\Psi(y)\right] = \sum \phi\left[F_1(y \mid e_i)\right] Pr(e_i)$$

et

$$\phi\left[\Psi(y)\right] - \text{borne inf.} \sum \phi\left[F(y \mid e_i)\right] Pr(e_i) = 0 \qquad (4.16)$$

C.Q.F.D.

B/ SI LE GAIN D'INFORMATION ELASTIQUE EST NUL, ON NE PEUT AFFIRMER QUE X ET Y SONT EN CORRELATION DURE.

Un exemple suffira à nous montrer qu'il en est bien ainsi.

On prend la variance comme mesure de l'incertitude.

a) Supposons que X est une v.a. discontinue pouvant prendre les valeurs $x_1, x_2, \ldots x_n$ et soit $F_{x_i}(y)$ les fonctions de répartition liées en ces points (on a vu que ces fonctions de répartition existent au sens classique) dès lors d'après (4.5) et (4.6), on aura :

$$\Psi(y) = \sum F_{x_i}\left[y + \omega(x_i)\right] Pr(X = x_i) \qquad (4.17)$$

et la v.a. $\mathcal{Y}$ qui a pour fonction de répartition $\Psi(y)$ aura une variance $\sigma^2_{\mathcal{Y}}$ et une moyenne $\overline{\mathcal{Y}}$.

et on aura, si on prend la variance comme mesure de l'incertitude

$$\left.\begin{aligned} & E = \sigma^2 - \text{borne inf.} \sum \sigma^2_{e_k} Pr(e_k) \\ & \text{avec } \sum e_k = R_1 \end{aligned}\right\} \qquad (4.18)$$

où $\sigma^2_{e_k}$ désigne toujours la variance de la v.a. Y_{e_k} de fonction de répartition $F(y \mid e_k)$

Or, d'après la formule (3.6), si $e_k = e_j + e_{j'}$ on aura :

$$\sigma^2_{e_k} Pr(e_k) \geqslant \sigma^2_{e_j} Pr(e_j) + \sigma^2_{e_{j'}} Pr\, e_{j'}$$

On en déduit aisément que :

$$\text{borne inf.} \sum_k \sigma^2_{e_k} Pr(e_k) = \sum_1^n \sigma^2_i Pr(X = x_i)$$

où σ^2_i désigne l'écart type de la v.a. de fonction de répartition $F_{x_i}(y)$ de sorte que (4.18) s'écrit :

$$E = \int_{-\infty}^{+\infty} (y - \overline{\mathcal{Y}})^2 d \sum_1^n F_{x_i}\left[y + \omega(x_i)\right] Pr(X=x_i) - \sum_1^n \sigma^2_i Pr(X=x_i)$$

$$= \sum_1^n \int_{-\infty}^{+\infty} \left[\eta - \omega(x_i) - \overline{\mathcal{Y}}\right]^2 dF_{x_i}(\eta) - \sum_{i=1}^n \sigma^2_i Pr(X=x_i) \qquad (4.19)$$

Or nous aurons :

$$\int_{-\infty}^{+\infty} \left[\eta - \omega(x_i) - \overline{\mathcal{Y}}\right]^2 dF_{x_i}(\eta) \geqslant \sigma^2_i \qquad (4.20)$$

le signe égal étant obtenu dans (4.20) si et seulement si :

$$\mathcal{Y} + \omega (x_i) = \int_{-\infty}^{+\infty} y \, d\, F_{x_i}(y)$$

On en déduit aisément que :

Théorème :

La condition nécessaire et suffisante pour que E = 0 quand on prend la variance comme mesure de l'incertitude et quand X est susceptible de prendre un nombre fini de valeurs $x_1, x_2 \dots x_n$ est que :

$$\omega (x_i) = K + \int y \, d\, F_{x_i} (y) \qquad (4.21)$$

où K est une constante arbitraire.

b) Cas où le couple X Y admet une densité de probabilité f(x,y).

Nous serons alors obligés évidemment de supposer $\varphi(x)$ mesurable au sens de Lebesgue.

Dans ces conditions, les intégrales $\int_{-\infty}^{+\infty} y^2 f(x,y)\, dy$ étant supposées toutes bornées, on voit aisément que l'on aura encore :

$$E = 0 \quad \text{si} \quad \varphi(x) = K + \frac{\int_{-\infty}^{+\infty} y\, f(x,y)\, dy}{\int_{-\infty}^{+\infty} f(x,y)\, dy}$$

en tous les points où le dénominateur n'est pas nul.

4 — GÉNÉRALISATION DES NOTIONS PRÉCÉDENTES

A/ RETOUR A LA NOTION DE CORRELATION DURE :

Il est aisé de démontrer que :

La condition nécessaire et suffisante pour que Y soit en corrélation dure, relativement à x (cf. définition p. 2.15) est qu'il existe une fonction univalente de x, $\varphi(x)$ telle que la v.a. $Y - \varphi(X)$ soit indépendante de X.

En effet, la condition nécessaire et suffisante pour qu'on puisse trouver pour le couple aléatoire (X, Y) un système admissible de fonctions de répartitions liées $\{ F_x(y) \}$ tel que $F_x(y) = C\,[\, y + \varphi(x)\,]$ est que l'on puisse trouver pour le couple $(Y - \varphi(X), X)$ un système admissible de fonctions de répartition liées $[\mathcal{F}_x(y)]$ telles que $\mathcal{F}_x(y) = C(y)$ et d'après le théorème de la p. 2.14, la condition nécessaire et suffisante pour qu'il en soit ainsi est que les v.a. $Y - \varphi(X)$ et X soient indépendantes.

Dire que l'on peut remplacer en gros la surface de probabilité considérée par une surface de probabilité en corrélation dure équivaut donc à dire que les v.a. $Y - \varphi(X)$ et Y sont grossièrement indépendantes.

La fonction de répartition de $Y - \omega(X)$ n'est autre que $\Psi(y)$ comme on le voit d'après (4.11). Il en résulte que :

$$D = \mathfrak{I}_Y - \mathfrak{I}_{Y-\omega(X)} \qquad (4.23)$$

$$E = \mathfrak{I}_{Y-\omega(X)} \quad - \text{borne inf.} \sum \phi[F(y \mid e_i)]\Pr(e_i) \qquad (4.24)$$

B/ GAIN D'INFORMATION RELATIF AU SCHEMA DE LA CORRELATION ISOGENE.

Le principal intérêt du schéma précédent est de nous permettre de construire très rapidement divers schémas simples, plus généraux que lui.

Ainsi la condition nécessaire et suffisante pour que Y soit en corrélation isogène relativement à x (cf. définition p. 2.17) est qu'il existe deux fonctions univalentes de x : $\varphi(x)$ et $\lambda(x)$ telles que la v.a. $\frac{Y - \varphi(X)}{\lambda(X)}$ soit indépendante de X.

Dès lors, étant données deux fonctions $\omega(x)$ et $\mu(x)$ on pourra définir des quantités $\mathcal{D}$ et $\mathcal{E}$ analogues aux quantités D et E définies précédemment. On aura :

$$\mathcal{D} = \mathfrak{I}_Y - \mathfrak{I}\ \frac{Y - \omega(X)}{\mu(X)} \tag{4.25}$$

$$\mathcal{E} = \mathfrak{I}\ \frac{Y - \omega(X)}{\mu(X)} - \text{borne inf.} \sum_i \phi\left[F(y \mid e_i)\right] \Pr(e_i) \tag{4.26}$$

on sera alors amené à poser :

$$\Psi^*(y) = P\left[Y \leqslant \mu(x)\left[\omega(x) + y\right]\right] \tag{4.27}$$

et :

$$F_1^*\left[y \mid e\right] = \frac{P\left[Y \leqslant \mu(x)\left[\omega(x) + y\right], X \subset e\right]}{P\left[X \subset e\right]}$$

si $P(X \subset e) \neq 0$. Mais ici on n'a plus $\mathcal{E} \geqslant 0$.

Il en résulte une sérieuse difficulté pour dire qu'un schéma de corrélation isogène est meilleur qu'un autre.

Certes, on peut toujours démontrer que :

<u>La condition nécessaire et suffisante pour que Y soit en corrélation isogène, relativement à x est qu'il existe 2 fonctions univalentes de x $\varphi(x)$ et $\lambda(x)$ telles que la v.a. $\frac{Y - \varphi(X)}{\lambda(X)}$, soit indépendante de X.</u>

et que par suite, si Y est en corrélation isogène par rapport à x, on pourra trouver un schéma probabiliste tel que :

$$\left.\begin{aligned} \mathcal{D} &= \Delta\mathfrak{I} \\ \mathcal{E} &= 0 \end{aligned}\right\} \tag{4.28}$$

Mais la réciproque n'est pas vraie. Et on pourra construire des schémas probabilistes qui donnent une idée tout à fait inexacte de F(x, y) et qui néanmoins sont tels que (4.28) soit vérifiée.

LECTURES RECOMMANDEES

Les personnes désireuses de connaitre les diverses généralisations de la loi de Bravais (définie dans Bravais [1]) liront avec fruit Bernstein [1]. Sarmanov [1] Prétorius [1] Narumi [1] Risser [1] .

CHAPITRE V

DE LA REGRESSION

1. — NOTION DE LIGNE DE REGRESSION

A/ LA DEFINITION CLASSIQUE DE LA LIGNE DE REGRESSION - SES GENERALISATIONS.

1°) La ligne de régression des moyennes.

Généralement, on ne parle que de la ligne de régression de la moyenne de Y en x d'un couple aléatoire possédant une densité de probabilité $f(x,y) = a(x).\ \alpha_x(y)$ (5.1)

La ligne de régression des moyennes est alors la courbe :

$$\varphi(x) = \int_{-\infty}^{+\infty} y\, \alpha_x(y)\, dy \tag{5.2}$$

2°) Conception de Cramér.

Pour Cramér, la notion de valeur typique est quelque chose de vague. C'est une quantité qui, comme la moyenne, la médiane ou le mode, peut être considérée comme le "point central" de la distribution (Cramér [1] p. 178). Dès lors, pour lui, la ligne de régression de Y en x sera définie dans deux cas particuliers.

a) Quand le couple aléatoire XY possède une densité de probabilité f(x, y)

On considère alors la valeur typique $\mathscr{C}(Y_x)$ de la v.a. Y_x qui a pour densité de probabilité $\alpha_x(y)$ définie par (5.1). La ligne de régression de Y en x est alors donnée par la courbe :

$$\mu(x) = \mathscr{C}(Y_x) \tag{5.3}$$

b) Quand X est une variable discrète (cf. Cramér [1] p. 272).

La fonction de répartition liée au point x tel que $P(X = x_i) \neq 0$ est alors déterminée comme il est dit au chapitre II (p. 210) ; soit $F(y \mid X = x_i)$ cette fonction de répartition. Alors la valeur typique $\mathscr{C}(Y_{x_i})$ de la v.a. Y_{x_i} de fonction de répartition $F(y \mid X = x_i)$ sera bien déterminée. Cramér propose d'appeler alors ligne de régression, la courbe obtenue en joignant par des segments de droite les points consécutifs $(x_i, \mathscr{C}(Y_{x_i}))$ $(x_k, \mathscr{C}(Y_{x_k}))$.

3° Première généralisation de la notion de régression.

Nous avons déjà donné cette généralisation p. 134 , le progrès est double.

a) La notion de valeur typique est bien précisée, c'est la valeur typique d'ordre k de Fréchet.

b) La ligne de régression est bien définie dès qu'il existe un système de fonctions de répartition $\{F_x(y)\}$ admissible relativement à la valeur typique d'ordre k de Fréchet (autrement dit la fonction de répartition du couple aléatoire XY peut être absolument quelconque).

Pour distinguer les lignes de régression définies par ce procédé, de celles que nous allons définir maintenant, nous les nommerons anciennes lignes de régression.

B/ NOUVELLES LIGNES DE REGRESSION.

1) Leur détermination.

Nous avons vu au chapitre IV, qu'un schéma probabiliste pouvait être considéré comme d'autant meilleur que D est plus grand. On sera donc tout naturellement amené à choisir la courbe $\omega(x)$ de sorte que D soit le plus grand possible. Ceci étant, nous adopterons la définition suivante :

Définition :

Nous appellerons nouvelle ligne de régression relative à la mesure ϑ de l'incertitude, toute fonction $y = \omega(x)$ telle que le maximum D^* de D soit atteint.

Plus précisément, comme nous voulons que D ait un sens, nous serons amenés à rechercher les fonctions $\omega(x)$ mesurables L_2 telles que $\vartheta[\Psi(y)]$ soit minimum quand (cf. 4.11) :

$$\Psi(y) = P\left[Y \leq \omega(x) + y\right]$$

Il convient dès à présent de remarquer qu'en vertu de (1.2), si $y = \omega(x)$ est une nouvelle ligne de régression $y = \omega(x) + a$ en est une également. On aura donc toujours non pas une nouvelle ligne de régression unique, mais une famille de telle lignes. Il est d'ailleurs aisé de voir que l'on peut retrouver ainsi les lignes de régression classiques.

Exemple I.

Recherchons d'abord l'ensemble des nouvelles lignes de régression quand on prend la variance comme mesure de l'incertitude.

On aura alors à rechercher la fonction $\omega(x)$ mesurable L_2 et telle que :

$$\vartheta\left[\Psi(y)\right] = \int (y - \bar{y})^2 \, d\Psi(y) \text{ minimum} \tag{5.4}$$

Quand :

$$\bar{y} = \int_{-\infty}^{+\infty} y \, d\Psi(y) \tag{5.5}$$

où $\Psi(y)$ est donnée par 4.11.
Et comme $\omega(x)$ est supposée mesurable L_2, on aura $\{F_x(y)\}$ désignant un système admissible de fonctions de répartition liées :

$$\Psi(y) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{y} dF_x\left[y + \omega(x)\right] \right\} dA(x) \tag{5.6}$$

d'où :

$$\bar{y} = \int_{-\infty}^{+\infty} \left\{ \int y \, dF_x\left[y + \omega(x)\right] \right\} dA(x) \tag{5.7}$$

$$= \int \left\{ \int \left[u - \omega(x)\right] dF_x(u) \right\} dA(x)$$

$$= \overline{Y} - \overline{\omega(x)} \tag{5.8}$$

en posant :

$$\overline{Y} = \int y \, dF(\infty, y) \tag{5.9}$$

et :

$$\overline{\omega(x)} = \int \omega(x) \, dA(x) \tag{5.10}$$

(5.4) s'écrit donc :

$$\phi\left[\Psi(y)\right] = \int \left\{ \int (y - \overline{Y} + \overline{\omega(x)})^2 \, dF_x\left[y + \omega(x)\right] \right\} dA(x)$$

$$= \int \left\{ \int \left[u - \overline{Y} - \omega(x) + \overline{\omega}(x)\right]^2 dF_x(u) \right\} dA(x)$$

ou en posant :

$$\omega(x) - \overline{\omega}(x) = \chi(x) \tag{5.11}$$

$$\phi\left[\Psi(y)\right] = \int \left\{ \int \left[u - (\overline{Y} + \chi(x))\right]^2 dF_x(u) \right\} dA(x) \tag{5.12}$$

et il est clair que le minimum de $\phi\left[\Psi(y)\right]$ sera atteint si toutes les intégrales $\int \left\{ u - \left[\overline{Y} + \chi(x)\right] \right\}^2 d\,F_x(u)$ sont minimum, ce qui se produira si $\overline{Y} + \chi(x) = \int y \, dF_x(y)$ ou $\omega(x) = \int y \, dF_x(y) + C^{te}$ (5.13)

Il convient toutefois de remarquer que ce calcul ne sera valable que si on peut trouver un système admissible de fonctions de répartition liées tel que toutes les intégrales utilisées aient un sens, ce qui n'est pas du tout évident.

De plus, notons que si le minimum de (5.12) est atteint pour une fonction $\chi^{(x)}$, il le sera pour toute fonction $\chi_1^{(x)}$ qui ne diffère de $\chi(x)$ que pour un ensemble de valeurs de x de probabilité nulle.

La réciproque est d'ailleurs vraie si sur un ensemble T tel que $P(X \subset T) \neq 0$, on a $\chi_2(x) \neq \chi(x)$ alors le minimum de $\phi\left[\Psi(y)\right]$ ne sera pas atteint pour $\chi_2(x)$. On peut donc énoncer le théorème suivant :

Théorème :

Si l'ancienne ligne de régression des moyennes $\mu(x)$ existe et si, de plus, toutes les intégrales dont il a été question plus haut ont un sens, la famille des nouvelles lignes de régression relatives à la variance comprend les courbes définies par :

1°) Les fonctions $\omega_1(x) = \mu(x) + C^{te}$

2°) Les fonctions de $\omega_2(x)$ telles que $\omega_2(x)$ ne diffère d'un $\omega_1(x)$ donné sur un ensemble de probabilité nulle.

Exemple II.

Prenons comme mesure de l'incertitude le moment typique d'ordre k de Fréchet, on aura alors :

$$\phi\left[\Psi(y)\right] = \min_a \int |y-a|^k \, d\Psi(y)$$

$$= \min_a \int \left[\int |y-a-\omega(x)|^k d\,F_x(y)\right] dA(x) \tag{5.14}$$

et le minimum de $\phi\left[\Psi\ (y)\right]$ sera encore atteint quand le minimum de $\int |y-a-\omega(x)|^k \, dF_x(y)$ est atteint, autrement dit si $a+\omega(x)$ est l'une quelconque des valeurs typiques d'ordre k de la v.a. Y_x de fonction de répartition $F_x(y)$.

La famille des nouvelles lignes de régression comprendra donc (sous réserve de conditions d'existence) :

1°/ La famille des courbes $\omega_1(x)$ parallèles à toute ancienne ligne de régression $\mu(x)$

$$\omega_1(x) = \mu(x) + K$$

2°/ La famille des courbes $\omega_2(x)$ telles que $\omega_2(x)$ ne diffère d'un $\omega_1(x)$ que sur un ensemble de probabilité nulle.

L'application la plus importante de ce théorème se rencontre quand $k = 1$; $\mu(x)$ est alors une ligne de régression des médianes. La médiane n'étant pas nécessairement unique, il en sera de même de $\mu(x)$. Dans ce cas, on pourra prendre pour $\omega_1(x)$ une courbe parallèle à l'un quelconque des $\mu(x)$ possibles.

Remarque :

Il y a lieu à notre sens de nous féliciter du fait que l'on n'obtient pas une nouvelle ligne de régression unique mais une famille de telles nouvelles lignes de régression car ceci nous montre qu'un très important groupe de propriétés des anciennes lignes de régression appartient aussi à d'autres courbes (les nouvelles lignes de régression).

La principale différence entre les anciennes et les nouvelles lignes de régression tient au fait que les premières sont définies à l'aide de la donnée d'une caractéristique fonctionnelle de position (Les c.f. relatifs aux v.a. X et X + a doivent être Ξ et $\Xi + a$), les secondes à l'aide d'une incertitude. Il est toutefois également possible de définir les anciennes lignes de régression à l'aide de la donnée d'une incertitude.

On peut ainsi définir l'ancienne ligne de régression de la moyenne comme étant la courbe définie par la fonction $\mu^*(x)$ qui rend minimum :

$$\mathfrak{M}\left[Y - \mu(X)\right]^2 \tag{5.15}$$

Et plus généralement, l'ancienne ligne de régression de la valeur typique d'ordre $k > 1$ comme étant la courbe définie par la fonction $\mu^*(x)$ qui rend minimum :

$$\mathfrak{M}\left|Y - \mu(X)\right|^k$$

Information dure et élastique quand $\omega(x)$ est une nouvelle ligne de régression.

Il est clair que si on prend la variance comme mesure de l'incertitude, on a :

$$D^* = \Delta\mathfrak{I} \qquad E^* = 0 \tag{5.16}$$

plus généralement, cette propriété sera encore vraie si on prend la valeur typique d'ordre k comme mesure de l'incertitude. Les diverses intégrales étant supposées exister, on a en effet, en tenant compte de (5.15) :

$$\begin{aligned} E^* &= \mathfrak{I}_{Y-\varphi(X)} - \mathfrak{M}\,\mathfrak{I}_{Y_x} \\ &= \min_a \int \left[\int |y - (a+\varphi(x)|^k \, dF_x(y)\right] dA(x) \\ &= 0 \end{aligned}$$

Il n'en sera pas de même si ϕ est une fonctionnelle strictement concave (4.12) s'écrit en effet alors :

$$\phi\left[\Psi(y)\right] \geqslant \sum \phi\left[F_1 \ (y \mid e_i)\right] \ Pr(e_i) \qquad (5.18)$$

et l'on ne peut avoir le signe égal dans (5.18) que si : $F_1(y \mid e_i) = \Psi(y)$ pour tout e_i tel que $P(e_i) \neq 0$; si donc nous supposons que (4.13) a lieu, nous ne pourrons avoir $E^* = 0$ que si $Y - \omega(X)$ est indépendante de X, c'est-à-dire si Y est en corrélation dure par rapport à x.

2. — NOTION DE DROITE DE REGRESSION

A/ LA DEFINITION DE LA DROITE DE REGRESSION.

1 - Anciennes droites de régression.

Dans (5.15), nous avons pu définir une ligne de régression des moyennes quelconques. Mais on peut remarquer que dans (5.15) on peut imposer $\mu(x)$ d'être une courbe continue, d'être dérivable, d'être une courbe d'une forme analytique donnée, une parabole ou une droite par exemple. Dans ce dernier cas, on obtiendra pour $\mu^*(x)$ l'ancienne droite de régression ajustée par la méthode des moindres carrés.

On aura donc pour déterminer cette ancienne droite de régression à déterminer les valeurs a^* b^* de a et de b qui rendent minimum $\mathfrak{M}(Y - a - bX)^2$; la droite de régression ainsi obtenue sera :

$$\frac{y - \bar{Y}}{\sigma_Y} = \rho \frac{x - \bar{X}}{\sigma_X} \qquad (5.19)$$

où σ_X et σ_Y sont les écarts types de X et Y,
ρ leur coefficient de corrélation linéaire.

On peut évidemment généraliser cette notion de la même manière que la précédente et chercher les valeurs a^* et b^* de a et b telles que

$$\mathfrak{M}\left|Y - a - bX\right|^k \text{ soit minimum.}$$

En particulier, si k = 1, on obtient ainsi la droite des moindres écarts.

2 - Nouvelles droites de régression.

Nous pouvons, dans le schéma probabiliste considéré au chapitre IV, imposer à $\omega(x)$ d'être d'une forme analytique simple, par exemple d'être une droite.

Nous essaierons de déterminer l'ensemble des valeurs a^* b^* de a et b pour lesquelles le maximum D_1^* de D est atteint quand on astreint $\omega(x)$ à être de la forme $\omega(x) = ax + b$ et nous appellerons nouvelles droites de régression relative à la mesure ϕ de l'incertitude, celles des droites ax + b qui rendent D maximum.

Il est clair que (4.11) s'écrira :

$$\Psi(y) = P\left[Y \leqslant aX + b + y\right] \qquad (5.20)$$

et que D sera précisément obtenu quand $\phi\left[\Psi(y)\right]$ est minimum.

Or, en vertu de (1.2), il est clair que si a^* et b^* sont des valeurs de a et b telles que $\phi\left[\Psi(y)\right]$ est minimum, il en sera de même de a^* et $b^* + c$ et ceci quel que soit c, autrement dit, quel que soit $\mathcal{C}$, $a^* x + \mathcal{C}$ sera elle aussi une nouvelle droite de régression relativement à l'incertitude $\phi\left[F(y)\right]$.

La famille des nouvelles droites de régression coincide donc avec la famille des droites $(a^* x + \mathcal{C})$ parallèles aux droites $a^* x$ rendant D maximum quand on impose à $\omega(x)$ d'être de la forme $\omega(x) = ax$.

B/ EXEMPLES :

En particulier, on peut prendre pour $\phi\left[F(y)\right]$ un écart typique d'ordre k. On voit immédiatement que :

Si nous prenons la variance comme mesure de l'incertitude, les nouvelles droites de régression sont les parallèles à la droite de régression classique (droite ajustée par la méthode des moindres carrés).

De même, si nous prenons l'écart moyen comme mesure de l'incertitude, les nouvelles droites de régression sont précisément les parallèles à la droite ajustée (ou aux droites ajustées) par la méthode des moindres écarts.

C/ DECOMPOSITION DU GAIN D'INFORMATION

On peut évidemment écrire :

$$\Delta \mathcal{I} = D_1^* + D^* - D_1^* + E^* \qquad (5.21)$$

et si (4.13) est vérifiée, les quantités D_1^*, $D^* - D_1^*$, E^* sont toutes trois non négatives.

La quantité $D^* - D_1^*$ mesure en particulier l'information gagnée en remplaçant le schéma probabiliste où $\omega(x)$ est assujetti à être une droite par un schéma probabiliste où $\omega(x)$ est quelconque (1).

Ainsi, le fait de remplacer le schéma probabiliste où $\omega(x)$ est une courbe quelconque par un schéma probabiliste où on impose à $\omega(x)$ d'être une droite, sera d'autant plus justifié que le rapport $\frac{D^* - D_1^*}{D^*}$ sera plus petit.

Si ce rapport n'est pas jugé assez petit, nous pourrons trouver un schéma mieux adapté de la manière suivante :

D/ REGRESSION PARABOLIQUE :

Supposons maintenant que nous imposions à $\omega(x)$ d'être un polynôme de degré n

$$\omega(x) = a_0 + a_1 x \quad + \quad + a_n x^n \qquad (5.22)$$

et que nous déterminions les valeurs a_0^*, $a_1^* \dots a_n^*$ de $a_0, a_1 \dots a_n$ telles que le maximum D_n^* de D soit atteint.

On voit comme précédemment que si :

$a_0^* + a_1^* x + \dots a_n^* x^n$ est une nouvelle parabole de régression de degré n, il en sera de même de :

$C + a_1^* x + \dots + a_n^* x^n$ et ceci quelle que soit la constante C.

Nous dirons que D_n^* est le gain d'information parabolique d'ordre n relatif à l'incertitude $\phi\left[F(y)\right]$

Il est clair que le gain d'information parabolique d'ordre m est plus petit ou égal au gain d'information parabolique d'ordre n si $m < n$

$$D_m^* \leqslant D_n^* \text{ si } m < n \qquad (5.23)$$

(1) Dans le cas particulier où on prend la variance comme mesure de l'incertitude on a $D^* - D_1^* = \mathcal{M}\left[\varphi(X) - \mathcal{M}\varphi(X) - aX\right]^2$ autrement dit, on peut trouver cette valeur, connaissant seulement $F(x,\infty)$ et les nouvelles lignes de régression.

On remarquera que (5.22) peut s'écrire :

$$\Delta \mathfrak{I} = D_1^* + D_2^* - D_1^* + \ldots + D_n^* - D_{n-1}^* + D^* - D_n^* + E^* \qquad (5.24)$$

Autrement dit, dans la pratique on pourra se faire plusieurs représentations de plus en plus précises de la surface de probabilité en prenant pour $\omega(x)$ des polynômes de degrés de plus en plus élevés. L'intérêt de la formule (5.24) est de nous indiquer à quel moment nous pourrons nous considérer comme satisfaits car nous pouvons à chaque fois évaluer à la fois l'information que nous avons effectivement gagnée $D_n^* - D_{n-1}^*$ et l'information qui reste à gagner $D^* - D_n^*$.

Nous pourrons par exemple nous arrêter quand $\frac{D^* - D_n^*}{D^*}$ est inférieur à un nombre donné. Ceci n'est d'ailleurs peut-être pas la seule considération qui doit entrer en ligne de compte, on peut aussi songer à s'arrêter simplement parce que $\frac{D_n^* - D_{n-1}^*}{D^*}$ est trop petit, etc...

Remarque :

Au lieu d'imposer à $\omega(x)$ d'être un polynôme, on peut évidemment lui imposer d'avoir une autre forma analytique simple donnée, dépendant d'un certain nombre de paramètres.

BIBLIOGRAPHIE

sur les définitions classiques de la notion de régression, voir :

Brambilla [1] Cramér [1] Pompilj [1] Salvemini [2] Jordan [4]

CHAPITRE VI

DE LA CORRÉLATION

Supposons que nous désirions nous faire une idée de la valeur que prendra un certain caractère aléatoire Y, par exemple la taille d'un garçon quand il aura atteint l'âge de 20 ans.

Nous pourrons avoir une idée a priori de cette taille par la fonction de répartition B(y) des tailles de tous les hommes de 20 ans. Mais nous nous ferons une idée meilleure de cette taille si nous connaissons la taille x du père du garçon et la fonction de répartition conditionnelle $F_x(y)$. Le problème que se pose le praticien est de juger de la valeur du renseignement supplémentaire ainsi obtenu. Ce problème est d'importance car au lieu de dépenser notre temps et notre argent à évaluer la taille du fils à l'aide de celle du père, nous pourrions l'utiliser à évaluer la taille du fils au moyen de celle de la mère, etc...

1. — INDICES DE CORRÉLATION

A) DEFINITION :

L'avantage que nous apportera la connaissance de la valeur prise par X sur la connaissance de la valeur que devra prendre la v.a. Y peut s'exprimer comme le rapport du gain d'information que nous avons effectivement réalisé à celui que nous pouvions espérer réaliser quand nous ne connaissions que la fonction de répartition de Y. C'est ce rapport que nous appellerons indice de corrélation C, on aura par définition :

$$C = \frac{\Delta \mathfrak{I}}{\mathrm{Max}\, \Delta \mathfrak{I}} \qquad (6.1)$$

où Max $\Delta \mathfrak{I}$ désigne le maximum que $\Delta \mathfrak{I}$ est susceptible d'avoir quand on sait seulement que Y a pour fonction de répartition B(y).

B) EXEMPLES :

1° On voit immédiatement que :

Si on prend la variance comme mesure de l'incertitude, nous obtenons pour C le carré η^2 du rapport de corrélation de Pearson toutes les fois que ce rapport est défini.

a) En effet, au sens classique η^2 n'est défini que si X est une v.a. discontinue susceptible de prendre les valeurs $x_1, \dots x_i, \dots x_m$ avec des probabilités positives $p_1, \dots p_m$ ($\sum_1^m p_i = 1$) et si de plus les moments liés $m_{x_i}^{(2)} = \int y^2 \, d F_{x_i}(y)$ existent : le carré du rapport de corrélation de Pearson sera alors : $\eta^2_{XY} = \dfrac{\sigma^2_Y - \sum p_i \sigma^2_i}{\sigma^2_Y}$ (6.2)

en posant :

$$\sigma^2_Y = \min_a \int (y-a)^2 \, d F(\infty, y)$$

$$\sigma^2_i = \min_{b_i} \int (y-b_i)^2 \, d F_{x_i}(y)$$

Et on reconnait au numérateur la quantité $\Delta \mathfrak{I}$

De plus, la fonction de répartition de Y étant donnée, on doit avoir max $\Delta \mathfrak{I} \leqslant \sigma^2_Y$ puisqu'une variance ne peut être négative ; mais quand Y est fonction univalente de x(cf. définition ch. II)on a $\sigma^2_i = 0$ donc $\sum p_i \sigma^2_i = 0$ il en résulte que max $\Delta \mathfrak{I} = \sigma^2_Y$.

b) Dans le cas où X n'est plus une v.a. discrète, nous continuerons à appeler carré du rapport de corrélation de Pearson la quantité $\eta^2 = \dfrac{\Delta \mathfrak{I}}{\text{Max } \Delta \mathfrak{I}}$ toutes les fois que cette quantité sera définie.

c) Le carré du rapport de corrélation de Pearson ainsi généralisé jouit d'importantes propriétés.

1° Si X et Y sont indépendantes, on a $\eta^2 = 0$ (c'est bien évident puisqu'alors $\Delta \mathfrak{I} = 0$.

2° Mais la réciproque n'est plus vraie. On peut même montrer que quand η^2 existe la condition nécessaire et suffisante pour que $\eta = 0$ est que $\int y \, d F(y \mid e)$ soit indépendante de e pour tout ensemble e pour lequel $F(y \mid e)$ est défini (cf. théorème p. 3, 8)

Si η^2 et l'ancienne ligne de régression des moyennes existent tous deux, il faut et il suffit pour que $\eta^2 = 0$ que l'ancienne ligne de régression des moyennes soit parallèle à Ox (1).

3° Si Y est fonction univalente de x en moyenne quadratique au sens strict, alors $\eta^2 = 1$.

En effet, d'après la fin du chapitre III, le numérateur de η^2 est nécessairement alors égal à σ^2_Y.

4° Réciproquement, si $\eta^2 = 1$, Y est presque certainement fonction univalente de x en moyenne quadratique au sens large. (cf. démonstration chapitre III).

On remarquera qu'avec la définition classique de η^2 dans laquelle on suppose que X ne peut prendre qu'un nombre fini de valeurs x_i avec des probabilités positives p, les conditions précédentes peuvent se simplifier. En posant $F_{x_i}(y) = \Pr(Y \leqslant y \mid X = x)$ on a en effet toutes les fois que η^2 est défini.

(1) On dit alors que Y est indépendante de X en moyenne.

1) 2) La condition nécessaire et suffisante pour que $\eta^2 = 0$ est que la moyenne liée $\overline{Y}_{x_i} = \int y \, d F_{x_i}(y)$ soit constante (1).

3) 4) La condition nécessaire et suffisante pour que $\eta^2 = 1$ est qu'il existe un nombre $\varphi(x_i)$ tel que :

$$F_{x_i}(y) = \begin{cases} 0 \text{ pour } y < \varphi(x_i) \\ 1 \text{ pour } y \geq \varphi(x_i) \end{cases}$$

2° Plus généralement, prenons l'écart typique d'ordre k de Fréchet comme mesure de l'incertitude, on obtient alors comme précédemment un indice $C^{(k)} = \frac{\Delta \mathfrak{I}}{\operatorname{Max} \Delta \mathfrak{I}}$ jouissant des propriétés suivantes :

1) Si X et Y sont indépendantes : $C^{(k)} = 0$

2) La condition nécessaire et suffisante pour que $C^{(k)} = 0$ est qu'on puisse trouver une quantité a indépendante de e et telle que le minimum par rapport à b de $\int |y-b|^k \, dF(y \mid e)$ soit obtenu pour b = a et ceci pour tout ensemble e pour lequel $F(y \mid e)$ est défini. En d'autres termes que les v.a. Y_e de f. de r. $F(y \mid e)$ aient toutes une valeur typique d'ordre k commune.

3) Si Y est fonction univalente de x en moyenne d'ordre k au sens strict, alors $C^{(k)} = 1$. (1)

4) Si $C^{(k)} = 1$, Y est fonction univalente de x en moyenne d'ordre k au sens large.

En particulier, si $k > 1$ et si X est une v.a. pouvant prendre un nombre fini de valeurs x_i avec des probabilités positives p_i.

1) 2) La condition nécessaire et suffisante pour que $C^{(k)} = 0$ est que la valeur typique liée d'ordre k de Fréchet soit constante (On appelle valeur typique liée d'ordre k de Fréchet au point x_i celle des quantités a_i qui rend minimum l'intégrale $\int |y - a_i|^k \, d F_{x_i}(y)$.

3) 4) La condition nécessaire et suffisante pour que $C^{(k)} = 1$ est qu'il existe un nombre $\varphi(x_i)$ tel que :

$$F_{x_i}(y) = \begin{cases} 0 \text{ pour } y < \varphi(x_i) \\ 1 \text{ pour } y \geq \varphi(x_i) \end{cases}$$

Si k = 1 les conditions 3) 4) restent inchangées mais on doit légèrement modifier les conditions 1) 2) pour tenir compte du fait qu'on n'a plus nécessairement une seule valeur typique; la condition 1) 2) deviendra alors :

1) 2) La condition nécessaire et suffisante pour que $C^{(1)} = 0$ est qu'on puisse trouver une quantité a, égale à une médiane liée pour tout x_i (on appelle médiane liée à x_i toute médiane de la v a Y_{x_i} de f. de r. $F_{x_i}(y)$).

3° Supposons enfin que les v.a. X et Y sont totalement discontinues et susceptibles de prendre seulement les valeurs x_i, y_j et que nous prenons l'entropie comme mesure de l'incertitude.

Soient :

p_{ij} la probabilité du couple x_i, y_j

(1) Ceci tient somme toute uniquement au fait que, si on prend le moment typique d'ordre k de Fréchet pour une mesure de l'incertitude la condition V p. 120 est elle aussi vérifiée.

$p_{i.}$ la probabilité marginale de $X = x_i$ $p_{i.} = \sum_j p_{ij}$

$p_{.j}$ la probabilité marginale de $Y = y_j$ $p_{.j} = \sum_i p_{ij}$

$p_{j|i}$ la probabilité conditionnelle de y_j quand on sait que $X = x_i$ $\quad p_{j|i} = \dfrac{p_{ij}}{p_{i.}}$

on a :

$$p_{j|i} \geqslant 0 \qquad \sum_j p_{j|i} = 1$$

Si nous prenons l'entropie comme mesure de l'incertitude, nous aurons :

$$\Delta \mathfrak{I} = - \sum p_{.j} \, L \, p_{.j} + \sum_i p_{i.} \sum_j p_{j|i} \; L \, p_{j|i} \tag{6.5}$$

Comme on a $0 \leqslant p_{j|i} \leqslant 1$ on a $p_{j|i} \, L \, p_{j|i} \leqslant 0$ et le maximum de cette quantité sera précisément égal à zéro et sera obtenu quand tous les $p_{j|i}$ sont nuls à l'exception d'un seul qui est égal à 1. On aura donc :

$$C = \frac{\Delta \mathfrak{I}}{\text{Max.}\, \Delta \mathfrak{I}} = \frac{- \sum_j p_{.j} \, L \, p_{.j} + \sum_i p_{i.} \sum_j p_{j|i} \; L \, p_{j|i}}{- \sum_j p_{.j} \; L \, p_{.j}}$$

Et en tenant compte des résultats du chapitre III, on voit que C est un indice qui vérifie les conditions de Fréchet, c'est-à-dire que c'est un nombre compris entre 0 et 1 et que l'on a :

1°) 2°) C = 0 si et seulement si X et Y sont indépendantes.

3°) 4°) C = 1 si et seulement si Y est fonction univalente de x (c'est-à-dire si pour toute valeur de i un seul des p_{ij} est différent de zéro).

2. — INDICES DE CORRELATION DURE

A) DEFINITION :

Mais nous avons vu que dans la pratique on pouvait se faire une idée moins précise, mais plus simple de la dépendance qui existe entre X et Y en remplaçant le couple aléatoire XY par un couple aléatoire $X^* Y^*$ tel que Y^* soit en corrélation dure relativement à x comme il est expliqué aux chapitres IV et V. Nous avons vu que nous pouvions faire cette substitution de différentes manières plus ou moins précises et que le gain d'information relatif à ces divers schémas probabilistes ne peut être supérieur à $\Delta \mathfrak{I}$

Ceci posé, il est utile d'avoir un indice capable de caractériser la valeur des schémas probabilistes considérés au chapitre V. La valeur d'un schéma donné ou indice de corrélation dure C_D s'exprimera comme le rapport de l'information gagnée par ce schéma à l'information qu'à priori nous pouvions espérer gagner connaissant simplement la fonction de répartition de Y.

$$C_D = \frac{D}{\text{Max.}\, \Delta \mathfrak{I}} \tag{6.7}$$

Parmi tous les schémas probabilistes possibles, nous nous bornerons dans la suite à ceux qui sont les mieux adaptés. Autrement dit, nous n'examinerons ici que le cas où :

1) D est le maximum D^* de D (cf p.156) quand la courbe $y = \omega(x)$ est quelconque.

2) D est le maximum D_1^* de D (cf.p. 159) quand la courbe $y = \omega(x)$ est une droite.

3) D est le maximum D_n^* de D (cf. p. 160) quand la courbe $y = \omega(x)$ est une parabole de degré n.

Nous désignerons les trois indices correspondants par :

$$C_{D^*} \quad , \quad C_{D_1^*} \quad , \quad C_{D_n^*}$$

En vertu de (5.23) on a évidemment :

$$C \geqslant C_{D^*} \qquad C_{D_n^*} \geqslant C_{D_1^*} \qquad (6.8)$$

D'autre part, les quantités D^*, D_n^*, D_1^* étant non négatives C, C_{D^*}, $C_{D_n^*}$, $C_{D_1^*}$ seront eux aussi non négatifs. Ceci posé, examinons ce que deviennent ces indices dans quelques cas particuliers .

B) EXEMPLES :

1° On prend la variance comme mesure de l'incertitude

a) la courbe $y = \omega(x)$ est supposée quelconque.

Nous avons vu au chapitre V que le maximum D^* de D était alors obtenu quand $y = \omega(x)$ est une ligne de régression des moyennes. Maintenant nous examinons uniquement l'avantage que nous retirons du schéma probabiliste considéré dans la définition des nouvelles lignes de régression.

Cet avantage s'exprime par :

$$C_{D^*} = \frac{D^*}{\text{Max.}\ \Delta \mathfrak{I}} \qquad (6.9)$$

où $\qquad D^* = \mathfrak{I}_Y$ - borne inf. $\sum \sigma^2_{e_i}\ \Pr(e_i)$

par conséquent C_{D^*} est bien défini et égal à η^2 toutes les fois que cette quantité est bien définie (cf. plus haut).

Autrement dit, quand on prend la variance comme mesure de l'incertitude, on a toujours $C = C_{D^*}$ (6.10)

b) la courbe $y = \omega(x)$ est supposée être une droite.

Le maximum D_1^* de D est alors atteint quand $y = \omega(x)$ est la droite de régression. Le couple aléatoire XY est maintenant considéré comme bien représenté par un couple aléatoire $X^* Y^*$ tel que Y^* est à la fois en corrélation dure et en régression linéaire par rapport à x. L'avantage que nous retirons de ce schéma probabiliste s'écrit :

$$C_{D_1^*} = \frac{D_1^*}{\text{Max}\ \Delta \mathfrak{I}} \qquad (6.11)$$

et un calcul simple montre que $C_{D_1^*}$ n'est autre que le carré du coefficient de corrélation linéaire $C_{D_1^*} = r^2$

c) la courbe $y = \omega(x)$ est supposée être une parabole de degré n.

Le maximum D_n^* de D est encore obtenu quand $\omega(x)$ est la parabole de régression. On a :

$$C_{D_n^*} = \frac{D_n^*}{\text{Max.}\ \Delta \mathfrak{I}} \qquad (6.12)$$

On aura évidemment $C_{D_n^*} = 0$ si la parabole de régression de degré n se réduit à une droite parallèle à l'axe des x (l'école italienne dit alors qu'on a indépendance parabolique d'ordre n cf. Salvemini [3] ou Pompilj [1]).

On aura $C_{D_n^*} = 1$ si et seulement si Y est presque sûrement fonction univalente de x : $y = \varphi(x)$ et si de plus $\varphi(x)$ est une parabole de degré n au plus.

2°) On prend la valeur typique d'ordre k comme mesure de l'incertitude.

D^*, D_1^*, D_n^* étant toujours définies comme il est dit au chapitre V, on posera encore :

$$C_{D^*}^{(k)} = \frac{D^*}{\text{Max } \Delta \mathfrak{I}}, \quad C_{D_1^*}^{(k)} = \frac{D_1^*}{\text{Max } \Delta \mathfrak{I}}, \quad C_{D_n^*}^{(k)} = \frac{D_n^*}{\text{Max } \Delta \mathfrak{I}} \qquad (6.13)$$

et l'on aura encore $C^{(k)}$ étant défini par (6.4) :

$$C^{(k)} = C_{D^*}^{(k)} \qquad (6.14)$$

$$1 \geqslant C^{(k)} \geqslant C_{D^*}^{(k)} \geqslant C_{D_n^*}^{(k)} \geqslant C_{D_1^*}^{(k)} \geqslant 0 \qquad (6.15)$$

et l'on aura encore $C_{D_n^*}^{(k)} = 0$ si et seulement si on peut trouver une nouvelle parabole de régression de degrés n parallèle à l'axe 0x.

De même, on aura encore $C_{D_n^*}^{(k)} = 1$ si et seulement s'il existe une parabole de degrés n : $\varphi(x) = a_0 + a_1 x \ldots + a_n x^n$ et un système admissible de fonctions de répartition $F_x(y)$ tels que :

$$F_x(y) = \begin{vmatrix} 0 \text{ pour } y < \varphi(x) \\ 1 \text{ pour } y \geqslant \varphi(x) \end{vmatrix}$$

NOTE HISTORIQUE

Le rapport de corrélation de Pearson a été défini dans Karl Pearson [3]

CHAPITRE VII

INCERTITUDES, INDICES DE CORRÉLATION ET LIGNES DE RÉGRESSION GÉNÉRALISÉES POUR UN COUPLE DE NOMBRES ALÉATOIRES

1. — GÉNÉRALISATION DE LA NOTION D'INCERTITUDE

Nous proposons dans ce chapitre en quelque sorte de généraliser la notion d'incertitude d'une manière analogue à celle que l'on emploie en algèbre pour généraliser la notion de distance. Le fait que l'on puisse découvrir des propriétés intéressantes de l'espace en ne considérant que la notion d'écart de Fréchet entre 2 points nous parait à ce propos particulièrement remarquable. Par analogie, nous attacherons à tout couple de fonctions de répartition $F(y)$, $G(y)$ un nombre réel positif négatif ou nul $\Omega\left[F(y), G(y)\right]$

$\Omega\left[F(y), G(y)\right]$ sera en quelque sorte un écart algébrique entre F et G et nous lui imposerons évidemment d'être nul si $F(y) \equiv G(y)$.

$$\Omega\left[F(y), G(y)\right] = 0 \quad \text{si} \quad F(y) \equiv G(y) \qquad (7.1)$$

Mais nous serons aussi dans la suite conduits à imposer à $\Omega\left[F(y), G(y)\right]$ la condition (7.2) analogue à (1.2).

Nous aurons :

$$\frac{1}{2}\left\{\Omega\left[G(y), F_1(y)\right] + \Omega\left[G(y), F_2(y)\right]\right\} \geqslant \Omega\left[G(y), \frac{F_1(y)+F_2(y)}{2}\right] \qquad (7.2)$$

Nous supposerons donc que (7.2) est vérifiée dans un certain domaine. Plus précisément, nous supposerons seulement cf (3.9) que (7.2) est vérifiée toutes les fois que $F_1(y)$, $F_2(y)$, $G(y)$ sont des $F_k(y)$ c'est-à-dire sont de la forme $\sum_i \lambda_i F(y \mid e_i) \quad (\sum \lambda_i = 1)$

Nous supposerons en outre que quels que soient les $F_k(y)$ les $\Omega\left[G(y), F_k(y)\right]$ restent compris entre deux nombres donnés m et M.

Avec ces deux seules hypothèses on pourra démontrer comme au chapitre III que quels que soient les P_i positifs rationnels ou non tels que $\sum p_i = 1$ on aura :

$$\sum p_i\, \Omega\left[G(y), F_i(y)\right] \geqslant \Omega\left[G(y), \sum p_i F_i(y)\right] \qquad (7.3)$$

ou

$$\sum \mathrm{Pr}(e_i)\ \Omega\left[G(y), F(y \mid e_i)\right] \geqslant \Omega\left[G(y), \sum F(y \mid e_i)\ \mathrm{Pr}(e_i)\right] \quad (7.4)$$

Si nous faisons $G(y) = B(y)$ compte tenu (7.1), (7.4) s'écrit en remarquant que $\sum F(y \mid e_i)\ \mathrm{Pr}(e_i) = B(y)$

$$\sum \mathrm{Pr}(e_i)\ \Omega\left[B(y), F(y \mid e_i)\right] \geqslant 0 \quad (7.5)$$

et l'on aura évidemment dans (7.5) le signe égal si quelque soit e_i

$$F(y \mid e_i) \equiv B(y)$$

De plus, si les $e_i^{(n)}$ sont les mailles du réseau d'ordre n, il est clair que :

$$\sum \mathrm{Pr}(e_i^{(n)})\ \Omega\left\{B(y), F\left[y \mid e_i^{(n)}\right]\right\} \leqslant \sum \mathrm{Pr}(e_i^{(n+1)})\ \Omega\left\{B(y), F\left[y \mid e_i^{(n+1)}\right]\right\}$$

2. — GÉNÉRALISATION DE LA NOTION DE GAIN D'INFORMATION

A) DEFINITION :

Première généralisation :

$\Omega\left[G(y), F(y)\right]$ étant supposé choisi comme nous l'avons dit précédemment, nous appellerons gain d'information généralisé la quantité :

$$\Delta \mathfrak{I} = \text{borne sup.} \sum \mathrm{Pr}(e_i)\ \Omega\left[B(y), F(y \mid e_i)\right] \quad (7.6)$$

Il est clair qu'on aura toujours $\Delta \mathfrak{I} \geqslant 0$ (7.7)
et qu'on aura nécessairement le signe égal dans (7.7) si X et Y sont indépendantes, mais que la réciproque n'est pas nécessairement vraie.

Deuxième généralisation :

Nous prenons simplement (7.6) comme définition du gain d'information mais sans imposer Ω de vérifier nécessairement les conditions qui ont servi à prouver (7.7). Nous imposons uniquement à Ω d'être tel que (7.7) soit vérifiée pour tout couple aléatoire XY pour lequel $\Delta \mathfrak{I}$ est défini et de plus d'être tel que $\Delta \mathfrak{I} = 0$ si X et Y sont indépendantes.

B) EXEMPLES :

Les conditions précédentes seront en particulier vérifiées dans les 3 cas particuliers suivants :

1) $\Omega\left[F(y), G(y)\right]$ est une fonction concave du vecteur $F(y) - G(y)$ autrement dit si on pose $F_1(y) - G(y) = z_1$, $F_2(y) - G(y) = z_2$
$\Omega\left[F_1(y), G(y)\right] = \omega\ (z_1)$ on a

$$\frac{1}{2}\left[\omega\ (z_1) + \omega\ (z_2)\right] \geqslant \omega\left[\frac{z_1+z_2}{2}\right]$$

2) $$\Omega\left[F(y), G(y)\right] = \phi\left[F(y)\right] - \phi\left[G(y)\right] \quad (7.8)$$

où ϕ est une incertitude. Dans ce cas, le fait que $\Delta \mathfrak{I} = 0$ entrainera l'indépendance de X et Y si nous supposons que ϕ est une fonctionnelle strictement concave, mais il n'en sera pas nécessairement de même si nous supposons seulement que ϕ est une fonctionnelle concave.

3) $\Omega\left[F(y), G(y)\right]$ est une mesure de l'écart de Fréchet (1) entre les deux lois de probabilité F(y), G(y) et on aura toujours $\Omega\left[F(y), G(y)\right] \geqslant 0$ le signe égal ne pouvant être obtenu que si $F(y) \equiv G(y)$.

Dans ce cas, on aura $\Delta\mathfrak{I} = 0$ si et seulement si X et Y sont indépendantes. Dans la suite, nous serons amenés à considérer comme cas particulièrement intéressant le cas où l'aire (prise en valeur absolue) comprise entre les courbes F(y) et G(y) est prise comme mesure de l'écart entre F(y) et G(y) (2)

$$\Omega\left[F(y), \quad G(y)\right] = \int \left| F(y) - G(y) \right| dy \tag{7.10}$$

Dans le cas où les v.a. Y_1 et Y_2 de fonctions de répartition F(y), G(y) ne peuvent prendre qu'un nombre fini de valeurs y_j avec les probabilités respectives $p_j^{(1)}$ $p_j^{(2)}$, nous serons également amenés à considérer les définitions suivantes de la distance :

$$\Omega\left[F(y), G(y)\right] = \sum_j \left| p_j^{(1)} - p_j^{(2)}\right| \tag{7.11}$$

$$\Omega\left[F(y), G(y)\right] = \sum_j \left| p_j^{(1)} - p_j^{(2)}\right|^2 \tag{7.12}$$

Toutes ces définitions d'ailleurs ne s'imposent nullement. Nous n'attirons l'attention sur elles que parce qu'elles nous permettront tout à l'heure de retrouver divers indices de corrélation qui apparaissent d'un grand intérêt théorique.

3. — GÉNÉRALISATION DE LA NOTION D'INDICE DE CORRELATION

1°) L'indice de corrélation généralisé C sera toujours donné par la formule (6.1) $C = \frac{\Delta\mathfrak{I}}{\text{Max } \Delta\mathfrak{I}}$, dans laquelle $\Delta\mathfrak{I}$ désigne maintenant le gain d'information généralisé. Cette quantité ne sera considérée comme définie que si $\Delta\mathfrak{I}$ et Max $\Delta\mathfrak{I}$ existent et sont finis.

1) Nous aurons évidemment toujours C = 0 si X et Y sont indépendantes et :

2) Nous voyons immédiatement que si $\Omega\left[F(y), G(y)\right]$ vérifie (7.9) on ne pourra avoir C = 0 que si X et Y sont indépendantes.

3) Par contre, nous ne savons pas a priori si nous aurons C = 1 quand Y est fonction univalente de x en un des sens stochastiques définis dans les p. 144 et suivantes. Il n'en sera ainsi que si le maximum de $\Delta\mathfrak{I}$ est obtenu précisément quand Y est fonction univalente de x en ce même sens.

---

(1) Rappelons qu'on appelle espace écartisé de Fréchet ou espace E tout espace où on peut définir la convergence ou la limite au moyen d'un écart c'est-à-dire où
1° à tout couple de points a b de l'espace considéré correspond un nombre $(a,b)=(b,a) \geqslant 0$ nombre qu'on appellera écart de a et de b.
2° On a $(a,b) = 0$ si a et b ne sont pas distincts et dans ce cas seulement
3° Pour qu'une suite infinie de points $a_1, a_2, \ldots a_n \ldots$ soit convergente et ait pour limite le point a il faut et il suffit que l'écart (a_n, a) entre a_n et a tende vers 0 quand n croit indéfiniment.
Cette définition est plus générale que celle de Bourbaki puisqu'elle ne suppose pas l'inégalité triangulaire.

(2) Cet écart vérifiant l'inégalité triangulaire est une distance au sens de Fréchet - Par contre ce n'est pas une distance au sens de Bourbaki, car cet $\Omega\left[F(y), G(g)\right]$ n'est pas nécessairement fini pour tout couple de fonctions de répartition.

4) De même, nous ne pourrons avoir C = 1 que si Y est fonction univalente de x en un certain sens stochastique, seulement si le maximum de $\Delta \mathfrak{J}$ ne peut être obtenu que si Y est fonction univalente de x à ce sens stochastique.

2°) Les principaux indices de corrélation (1) considérés jusqu'à présent n'ont, à notre connaissance, été définis que si X est une v.a. discrète ne pouvant prendre que les valeurs x_i avec des probabilités p_i. Ce sont :

a) Si Y est une v.a. discrète les indices simples et quadratiques de connexion partielle de Gini.

En effet, soit p_{ij} la probabilité du couple x_i, y_j posons :

$$p_{.j} = \sum_i p_{ij} \qquad \text{et} \qquad p_{i.} = \sum_j p_{ij} \tag{7.13}$$

On retrouve l'indice simple de connexion partielle de Gini :

$$\beta_x = \frac{\sum_i \sum_j \left| p_{ij} - p_{i.} \, p_{.j} \right|}{2 \, (1 - \sum p^2_{.j})} \tag{7.14}$$

si $\Omega \left[F(y), G(y) \right]$ est donné par (7.11) c'est-à-dire

si $\Omega \left[F(y), F_{x_i}(y) \right] = \sum \left| \frac{p_{ij}}{p_{i.}} - p_{.j} \right|$

De même, on retrouve l'indice quadratique de connexion partielle de Gini :

$$C_x^2 = \frac{\sum_i \sum_j \frac{1}{p_{i.}} (p_{ij} - p_{i.} \, p_{.j})^2}{1 - \sum p_{.j}^2} \tag{7.15}$$

quand $\Omega \left[F(y), G(y) \right]$ est défini par (7.12) c'est-à-dire

si $\Omega \left[B(y), F_{x_i}(y) \right] = \sum_j \left(\frac{p_{ij}}{p_{i.}} - p_{.j} \right)^2$

b) Si Y est une v.a. quelconque on retrouve l'indice simple de dissemblance de Gini :

$$g = \frac{\int \int \left| F_x(y) - B(y) \right| dy \, dF(x, \infty)}{\int B(y) \left[1 - B(y) \right] dy} \tag{7.16}$$

si $\Omega \left[F(y), G(y) \right]$ est défini par (7.10)

Il convient de remarquer que les indices de Jordan, Mme Geiringer, Fréchet ne sont pas des indices de corrélation à notre sens nouveau.

---

(1) Nous conformant à la terminologie franco-anglaise, nous appelons indice de corrélation ce que les Italiens appellent "indice de connessone". Ainsi nous continuerons à dire rapport de corrélation de Pearson... bien que pour l'école italienne il s'agisse là d'un indice de connexion. Nous ne gardons le mot connexion que pour les indices découverts par l'école italienne.

4. — RETOUR A LA NOTION D'INFORMATION ÉLASTIQUE

Ici la généralisation des notions de gain d'information dure et élastique est relativement complexe et il est probable qu'en raison de cette complexité même cette généralisation sera relativement peu utile. Aussi nous bornerons-nous à considérer uniquement un cas particulièrement simple: celui où $\Omega\left[F(y), G(y)\right]$ est un écart, c'est-à-dire que nous supposons que :

$$\left.\begin{array}{l} \Omega\left[F(y),\ G(y)\right] \geqslant 0 \\ \Omega\left[F(y),\ G(y)\right] = 0 \ \text{si et seulement si } F(y) \equiv G(y) \end{array}\right\} \quad (7.17)$$

On sera donc amené à poser (voir ch. IV)

$$E = \text{borne sup.} \sum \Omega\left[\Psi\,(y), F(y \mid e_i)\right] Pr(e_i) \quad (7.18)$$

et nous aurons toujours $E \geqslant 0$.

Et nous serons amenés à considérer que le schéma est le mieux adapté possible si E est le plus petit possible.

Or, la valeur de la quantité E dépend uniquement de

$$\Psi\,(y) = P\left[Y \leqslant \omega\,(x) + y\right] \qquad (\text{cf } 4.11)$$

donc E est une fonctionnelle de ω (x)

Nous sommes donc tout naturellement conduits à appeler nouvelle ligne de régression relative à Ω toute fonction ω^* (x) telle que le minimum E^* de E soit atteint.

Il est bien clair que si nous prenons pour ω (x) la droite y = 0, nous trouverons $E = \Delta \mathfrak{I}$

Nous aurons donc nécessairement :

$$0 \leqslant E^* \leqslant \Delta \mathfrak{I}$$

Il est d'ailleurs clair qu'on aura $E^* = 0$ si et seulement si Y est en corrélation dure relativement à x.

De même nous pourrons définir les nouvelles droites de régression relatives à Ω, comme étant celles des droites ax + b pour lesquelles le minimum E_1^* de E est atteint quand on impose à ω (x) d'être une droite. Et nous aurons évidemment :

$$0 \leqslant E^* \leqslant E_1^* \leqslant \Delta \mathfrak{I}$$

VALEUR D'UN SCHEMA DE CORRELATION ISOGENE

Nous avons vu p. (4.10) qu'avec la définition classique de l'incertitude, nous ne pouvions juger aisément la valeur d'un schéma de corrélation isogène car nous n'avons plus nécessairement $\mathcal{E} \geqslant 0$. Au contraire, si Ω est donné par (7.17), nous serons conduits à poser :

$$\mathcal{E} = \text{borne sup.} \sum \Omega\left[\Psi\,(y),\ F(y \mid e_i)\right] Pr(e_i) \qquad (7.19)$$

avec :

$$\Psi\,(y) = P\left[Y \leqslant \lambda\,(x)\left[y + \varphi(x)\right]\right] \Big\}$$

nous aurons toujours :

$$\mathcal{E} \geqslant 0$$

Nous serons amenés à considérer le schéma comme le mieux adapté possible si on prend pour $\lambda(x)$ et $\varphi(x)$ des valeurs $\lambda^*(x)$ et $\varphi^*(x)$ telles que le minimum $\mathcal{E}^*$ de $\mathcal{E}$ soit atteint.

On aura évidemment :

$$0 \leqslant \mathcal{E}^* \leqslant \Delta \mathcal{J} \tag{7.20}$$

Il est évidemment également possible de considérer d'autres familles de schémas probabilistes et de déterminer par ce procédé le mieux adapté les schémas de la famille considérée.

On peut également comparer les schémas de 2 familles différentes par exemple, on peut comparer les 2 familles de schémas probabilistes considérés ici. La seconde étant incluse dans la première, on aura toujours

$$\mathcal{E}^* \leqslant E^*$$

LECTURES RECOMMANDEES

Les divers indices de corrélation dont il est question ici sont définis dans Brambilla [1] la lecture de Gini [1] (où est défini g) est tout particulièrement recommandée. - Après sa lecture il n'est pas du tout évident que g peut se mettre sous la forme (7.16). On s'en rendra compte toutefois en consultant Salvemini [1] la démonstration de ces auteurs a été reprise en français dans Fréchet [2] et dans Féron [1] pour les questions de Topologie voir Fréchet [16] et Bourbaki [1] .

CHAPITRE VIII

INFORMATION, REGRESSION ET CORRELATION DANS L'ESPACE EUCLIDIEN A n DIMENSIONS

Nous nous bornerons à examiner ce que deviennent les notions précédentes dans le cas simple souvent rencontré dans la pratique où l'on ne considère plus 2 variables aléatoires, mais n variables aléatoires $Z_1, Z_2 \ldots Z_n$ de fonction de répartition $F(z_1, z_2, \ldots z_n)$. Nous obtiendrons une généralisation des résultats précédents en faisant jouer aux p premières variables un rôle particulier par rapport aux $m = n-p$ suivantes. Autrement dit, les scalaires x, y considérés précédemment seront remplacés par des vecteurs $\vec{x}$ et $\vec{y}$ où $\vec{x}$ est un vecteur à p dimensions et $\vec{y}$ un vecteur à m dimensions.

Ceci posé, nous nous proposons de généraliser à l'espace les notions d'incertitude, de gain d'information, de régression et de corrélation, telles qu'elles ont été définies dans les chapitres I à VI.

1. — GÉNÉRALISATION DE LA NOTION D'INCERTITUDE

A) DEFINITION

Il nous faut maintenant définir ce que nous entendons par incertitude sur un vecteur aléatoire $\vec{Y}$ de fonction de répartition :

$$F(\vec{y}) = F(y_1, y_2 \ldots y_m) \qquad (8.1)$$

A cet effet, nous considérons une fonctionnelle $\phi\left[F(\vec{y})\right]$ à laquelle nous imposerons une condition analogue à (1.1). Cette condition sera :

$$\phi\left[\frac{F_1(y_1, y_2 \ldots y_m) + F_2(y_1, y_2 \ldots y_m)}{2}\right] \geqslant \frac{1}{2}\left\{\phi\left[F_1(y_1, \ldots y_m)\right] + \phi\left[F_2(y_1, \ldots y_m)\right]\right\} \qquad (8.2)$$

Si (8.2) est vérifiée, nous dirons encore que ϕ est une <u>fonctionnelle concave</u>. Si de plus, on ne peut avoir le signe égal dans (8.2) que si $F_1(\vec{y}) \equiv F_2(\vec{y})$ nous dirons que ϕ est une fonctionnelle strictement concave.

De plus, dans les études qui touchent à la régression, il est utile d'imposer à ϕ une condition analogue à (1.2) ; on devra avoir, quel que soit le vecteur $\vec{a}$ $(a_1, \ldots a_m)$:

$$\phi\left[F(\vec{y} - \vec{a})\right] = \phi\left[F(\vec{y})\right] \qquad (8.3)$$

B) EXEMPLES :

1) $\Lambda\left[F(y)\right]$ désignant une fonctionnelle concave ordinaire, il est clair que la fonctionnelle :

$$\phi\left[F(\vec{y})\right] = \sum_{i=1}^{m} k_i \, \Lambda \left[F(\infty, \ldots \infty, y_i, \infty, \ldots \infty)\right] \qquad (8.4)$$

(où $F(\infty, \infty \ldots \infty, y_i, \infty \ldots \infty)$ désigne la fonction de répartition marginale de Y_i et les k_i des constantes positives arbitraires) sera concave (1).

2) Nous pouvons généraliser la notion d'entropie dans le cas où $\vec{Y}$ est un vecteur aléatoire totalement discontinu. Soit $p_{i_1}, i_2, \ldots i_m$ la probabilité pour que le vecteur aléatoire $\vec{Y}$ ait pour coordonnées $y_1^{i_1}, y_2^{i_2}, \ldots y_m^{i_m}$.

Nous appellerons encore entropie de ce vecteur aléatoire la fonctionnelle :

$$\phi\left[F(\vec{y})\right] = \sum_{i_1, i_2, \ldots, i_m} p_{i_1, i_2, \ldots i_m} \, L \, p_{i_1, i_2, \ldots i_m} \qquad (8.5)$$

Il est clair que si $F_1(\vec{y})$ et $F_2(\vec{y})$ sont les fonctions de répartition de deux vecteurs aléatoires $\vec{Y}_1$ et $\vec{Y}_2$ totalement discontinus, $\dfrac{F_1(\vec{y}) + F_2(\vec{y})}{2}$ sera aussi la fonction de répartition d'un vecteur aléatoire totalement discontinu et qu'on aura ϕ étant donné par (8.5) :

$$\phi\left[\frac{F_1(\vec{y}) + F_2(\vec{y})}{2}\right] \geqslant \frac{1}{2}\left\{\phi\left[F_1(\vec{y})\right] + \phi\left[F_2(\vec{y})\right]\right\} \qquad (8.6)$$

le signe égal ne pouvant être obtenu dans (8.6) que si $F_1(y)$ est identique à $F_2(y)$.

3) Supposons enfin que nous désirions définir d'une manière judicieuse l'incertitude sur le vecteur aléatoire à 2 dimensions $\vec{Y}$ de fonction de répartition $F(\vec{y}) = F(u, v)$.

Le problème étant relativement simple, nous pourrons choisir $\phi\left[F(\vec{y})\right]$ de manière à lui donner une signification physique simple. Nous sommes ainsi fort tentés de prendre pour $\phi\left[F(u,v)\right]$ le gain d'information $\Delta\mathfrak{I}_{uv}$ ou l'un des gains d'information dure D^* ou D_1^*.

Pour pouvoir le faire, il faudrait prouver que $\Delta\mathfrak{I}$, D^*, D_1^* sont des fonctionnelles concaves. Or, il n'en est généralement pas ainsi. Bien au contraire. Nous allons donner un exemple où $\Delta\mathfrak{I}$, D^* et D_1^* sont des fonctionnelles convexes, ce qui nous permettra de prendre $-\Delta\mathfrak{I}$, $-D^*$, $-D_1^*$ comme mesures de l'incertitude.

Supposons que pour la détermination de $\Delta\mathfrak{I}$, D^* et D_1^* nous ayons pris la variance comme mesure de l'incertitude, autrement dit :

$$\mathcal{D}^*\left[F(u,v)\right] = \int (v-\overline{V})^2 \, dF(\infty, v) - \int\int (v-\overline{V}_u)^2 \, dF_u(v) \, dF(u,\infty) \qquad (8.7)$$

$$\mathcal{D}_1^*\left[F(u,v)\right] = \int (v-\overline{V})^2 dF(\infty\,; v) - \min_a \int\int (v-au-\overline{V})^2 \, dF_u(v) \, dF(u,\infty) \qquad (8.8)$$

(1) Plus généralement si $\psi(\lambda_1, \lambda_2, \ldots \lambda_m)$ est une fonction concave et croissante de l'ensemble des variables numériques $\lambda_1, \lambda_2, \ldots, \lambda_m$ (cf. Rado 1) et si les Λ_i sont des fonctionnelles concaves alors
$\Phi\left[F(y)\right] = \psi\left\{\Lambda_1\left[F(y, \infty \ldots \infty)\right], \ldots\ldots, \Lambda_i\left[F(\infty, \infty\, y_i\, \infty \ldots \infty)\right], \ldots \Lambda_m\left[F(\infty, \ldots \infty, y_m)\right]\right\}$
est une fonctionnelle concave.

où :

$$\overline{V} = \int v\, dF(\infty, v)$$
$$\overline{V}_u = \int v\, dF_u(v)$$

Un calcul simple montre que si $\mathcal{D}^*$ et $\mathcal{D}_1^*$ sont donnés par (8.7) et (8.8), on a :

$$\mathcal{D}^*\left[\frac{F'(u,v)+F''(u,v)}{2}\right] \leqslant \frac{1}{2}\left\{ \mathcal{D}^*\left[F'(u,v)\right] + \mathcal{D}^*\left[F''(u,v)\right] \right\} \quad (8.9)$$

$$\mathcal{D}_1^*\left[\frac{F'(u,v)+F''(u,v)}{2}\right] \leqslant \frac{1}{2}\left\{ \mathcal{D}_1^*\left[F'(u,v)\right] + \mathcal{D}_1^*\left[F''(u,v)\right] \right\} \quad (8.10)$$

Donc $\mathcal{D}^*$ et $\mathcal{D}_1^*$ sont des fonctionnelles convexes et par conséquent $-\mathcal{D}^*$ et $-\mathcal{D}_1^*$ sont des fonctionnelles concaves et peuvent être prises comme mesure de l'incertitude.

2. — GÉNÉRALISATION DE LA NOTION DE FONCTION DE RÉPARTITION LIÉE

Etant donné un vecteur aléatoire $\vec{Z}\,(Z_1, Z_2, \dots Z_n)$ de fonction de répartition $F(z_1, z_2, \dots z_p, \dots z_n)$, nous appellerons système admissible de fonctions de répartition liées à l'ensemble des p premières variables tout système de fonctions de répartition $F_{z_1, z_2, \dots z_p}(z_{p+1}, \dots z_n)$ tel que :

$$F(z_1, z_2, \dots z_p, \dots z_n) = \int_{-\infty}^{z_1}\int_{-\infty}^{z_2}\dots\int_{-\infty}^{z_p} F_{\xi_1, \xi_2, \dots \xi_p}(z_{p+1}, \dots z_n)\; dF(\xi_1, \dots \xi_n, \infty, \infty) \quad (8.11)$$

et ceci quels que soient $z_1, z_2, \dots z_n$.
(8.11) peut d'ailleurs s'écrire dans le langage vectoriel condensé suivant :

$$F(\vec{x}, \vec{y}) = \int_{-\infty}^{\vec{x}} F_{\vec{\xi}}(\vec{y})\; dF(\vec{\xi}, \infty) \quad (8.12)$$

Et de même que précédemment on peut montrer qu'il existe toujours une infinité de systèmes admissibles de fonctions de répartition liées.

Dans ce cas, on dit encore (cf. Cramér [1] p. 160) que les vecteurs aléatoires $\vec{X}$ et $\vec{Y}$ sont indépendants si :

$$F(\vec{x}, \vec{y}) = F(\vec{x}, \vec{\infty}).\; F(\vec{\infty}, \vec{y}) \quad (8.13)$$

Avec cette définition, nous pouvons généraliser le théorème p. 134 et énoncer que :

Si deux vecteurs aléatoires sont indépendants, on peut trouver un système admissible de fonction de répartition liées $\left\{F_{\vec{x}}(\vec{y})\right\}$ pour lequel $F_{\vec{x}}(\vec{y}) = F(\vec{\infty}, \vec{y})$ quel que soit $\vec{x}$ et réciproquement.

De même, par analogie avec les définitions du chapitre II, nous dirons que le vecteur aléatoire $\vec{Y}$ est en corrélation dure relativement à $\vec{x}$ si on peut trouver un système admissible de fonctions de répartition liées

$F_{\vec{x}}(\vec{y})$ telles que, quel que soit le vecteur $\vec{x}$ à n-p dimensions, on puisse trouver un vecteur $\gamma(\vec{x})$ à p dimensions tel que :

$$F_{\vec{x}}(\vec{y}) = \Psi\left[\vec{y} + \overrightarrow{\gamma(\vec{x})}\right] \tag{8.14}$$

Nous pourrons également définir ce que nous appelons système admissible de fonctions de répartition relativement à la fonctionnelle ϕ comme à la p. 131, l'équation (2.13) deviendra alors :

$$\phi\left[\mathcal{F}_{\vec{x}}(\vec{y})\right] = \lim_{\delta(T_i)\to 0} \inf.\ \phi\left\{F\left[\vec{y} \,\middle|\, T_i\right]\right\} \tag{8.15}$$

où $\delta(T_i)$ désigne le diamètre d'un intervalle arbitraire contenant l'extrémité du vecteur $\vec{x}$.

3. — GÉNÉRALISATION DE LA NOTION DE GAIN D'INFORMATION

A) DEFINITION :

Le gain d'information sera :

$$\Delta\mathfrak{I} = \text{borne sup.}\left\{\phi\left[F(\infty, \vec{y})\right] - \sum \phi\left[F(\vec{y} \mid e_i)\right] \Pr(e_i)\right\} \tag{8.16}$$

les e_i étant des intervalles de l'espace R_p tels que $\sum e_i = R_p$ et on montre comme précédemment qu'on a toujours :

$$\Delta\mathfrak{I} \geqslant 0 \tag{8.17}$$

et on montrera comme précédemment que si X et Y sont indépendantes, on a toujours :

$$\Delta\mathfrak{I} = 0$$

La réciproque, au contraire, ne sera vraie que si ϕ est une fonctionnelle strictement concave (il en sera par exemple ainsi si ϕ est donnée par (8.5).)

Par contre, dans le cas particulier où $\vec{Y}$ est un vecteur à une dimension alors que $\vec{X}$ est un vecteur à p dimensions le fait que $\Delta\mathfrak{I} = 0$ quand on prend la variance comme mesure de l'incertitude signifie simplement que la moyenne liée est indépendante de $\vec{x}$ (c'est-à-dire que l'on peut trouver un système admissible de fonctions de répartition $F_{\vec{x}}(y)$ tel que :

$$\int y \, dF_{\vec{x}}(y) = \text{Cte}$$

4. — SUR UN SCHÉMA PROBABILISTE

A) LE SCHEMA PROBABILISTE

Comme précédemment, nous pourrons remplacer en première approximation les vecteurs aléatoires $\vec{X}$, $\vec{Y}$ par des vecteurs $\vec{X}^*$, $\vec{Y}^*$ qui sont en corrélation dure (cf. formule 8.14).

Le couple de vecteurs aléatoires $\vec{X}^*$ $\vec{Y}^*$ aura pour fonction de répartition $F^*(\vec{x}, \vec{y})$ telle que :

$$F^*(\vec{\infty}, \vec{y}) = F(\vec{\infty}, \vec{y}) \tag{8.17}$$

$$P^*\left[\vec{Y} < \overrightarrow{\omega(\vec{x})} + \vec{h}, \vec{X} \subset E\right] = P(\vec{X} \subset E)\,\Psi(\vec{h}) \qquad (8.18)$$

Ces conditions entrainent d'ailleurs :

$$\Psi(\vec{y}) = P\left[\vec{Y} \leqslant \overrightarrow{\omega(\vec{x})} + \vec{y}\right] \qquad (8.19)$$

B) INTERPRETATION GEOMETRIQUE DANS UN CAS PARTICULIEREMENT SIMPLE.

Considérons le cas particulièrement simple où $n=3$ et où il existe une densité de probabilité $f(z_1, z_2, z_3)$. Nous pouvons considérer la loi de probabilité de l'ensemble des variables aléatoires z_1, z_2, z_3 comme bien déterminée par la donnée d'une répartition de masse continue dans l'espace, la densité au point z_1, z_2, z_3 étant $f(z_1, z_2, z_3)$.

Toutefois, en première approximation, nous pourrons considérer que l'ensemble des v.a. Z_1, Z_2, Z_3 est suffisamment bien décrit en associant à chaque point z_1, z_2, z_3, une densité de probabilité $f^*(z_1, z_2, z_3)$ qui soit d'une forme analytique simple.

Mais dans ce cas particulier, deux schémas simples nettement différents se présentent suivant que l'on considère $\vec{y}$ comme un vecteur de R_1 ou de R_2.

1°/ $\vec{y}$ est un vecteur de R_1 -

Alors la distribution de densités $f^*(z_1, z_2, z_3)$ peut être considérée comme produite par une distribution de masse $\Psi(z_3)$ sur un fil élémentaire se déplaçant parallèlement à l'axe des z de manière à couper la surface :

$$z_3 = \overrightarrow{\omega(\vec{x})} = \omega(z_1, z_2) \qquad (8.20)$$

en un point fixe (les fils élémentaires étant bien entendu supposés répartis avec la densité $A(z_1, z_2) = \int f^*(z_1, z_2, z_3)\, dz_3$

2°/ $\vec{y}$ est un vecteur de R_2 -

Alors la distribution des densités $f^*(z_1, z_2, z_3)$ peut être considérée comme produite par une distribution de masse $\Psi(z_2, z_3)$ sur un plan élémentaire perpendiculaire à $0z_1$ et animée d'un mouvement de translation le long de la courbe que nous pouvons écrire en langage vectoriel

$$\vec{\zeta} = \overrightarrow{\omega(\vec{x})} \qquad (8.21)$$

ou en langage algébrique ordinaire :

$$\begin{aligned} z_2 &= \omega_1(z_1) \\ z_3 &= \omega_2(z_1) \end{aligned} \qquad (8.22)$$

(les plans élémentaires étant bien entendu supposés répartis avec la densité

$$A(z_1) = \iint f^*(z_1, z_2, z_3)\, dz_2\, dz_3).$$

Revenons maintenant au cas général et aux notations du paragraphe précédent.

C) GAIN D'INFORMATION DURE ET GAIN D'INFORMATION ELASTIQUE

Par analogie avec les définitions du chapitre IV, nous définirons le gain d'information dure par :

$$D = \phi\left[F(\vec{\infty}, \vec{y})\right] - \phi\left[\Psi(\vec{y})\right] \qquad (8.23)$$

et le gain d'information élastique par :

$$E = \phi\left[\Psi(\vec{y})\right] - \text{borne inf.} \sum_i \phi\left[F(\vec{y}) \mid e_i)\right] \Pr e_i \qquad (8.24)$$

et tous les théorèmes du ch. IV relatifs à ces quantités se généralisent aisément.

5. — GÉNÉRALISATION DE LA NOTION DE RÉGRESSION

A) GENERALISATION DE LA NOTION DE LIGNE DE REGRESSION :

Nous serons amenés à associer à tout vecteur $\vec{x}$ de l'espace euclidien à p dimensions un vecteur $\overrightarrow{\varphi(\vec{x})}$ de l'espace euclidien à m = n-p dimensions et appellerons variété de régression de $\vec{Y}$ en $\vec{x}$ (1) une quelconque variété :

$$\vec{\eta} = \overrightarrow{\omega(\vec{x})} \qquad (8.25)$$

pour laquelle le maximum D^* de D est atteint, c'est-à-dire pour laquelle $\phi\left[\Psi(\vec{y})\right]$ est minimum quand :

$$\Psi(\vec{y}) = P\left[\vec{Y} \leqslant \overrightarrow{\omega(\vec{x})} + \vec{y}\right]$$

en langage ordinaire (8.25) s'écrit évidemment :

$$\left.\begin{aligned} z_{p+1} &= \omega_1(z_1, z_2, \ldots z_p) \\ z_{p+2} &= \omega_2(z_1, z_2, \ldots z_p) \\ z_n &= \omega_m(z_1, z_2, \ldots z_p) \end{aligned}\right\} \qquad (8.26)$$

Dans la suite, nous supposerons que ϕ a été choisie de manière à vérifier l'équation :

$$\phi\left[F(\vec{y} - \vec{a})\right] = \phi\left[F(\vec{y})\right] \qquad (8.27)$$

qui généralise (1.2). Il en résulte que si la variété définie par (8.26) est une "variété de régression" quels que soient $a_1 \ldots a_m$, la variété :

$$\left.\begin{aligned} z_{p+1} &= \omega_1(z_1, z_2 \ldots z_p) + a_1 \\ z_n &= \omega_m(z_1, z_2 \ldots z_p) + a_m \end{aligned}\right\} \qquad (8.28)$$

sera aussi une "variété de régression".

Il est d'ailleurs clair que si $\vec{Y}$ est en corrélation dure par rapport à $\vec{x}$ cf. (8.14), quelle que soit la fonctionnelle $\phi\left[F(\vec{y})\right]$ vérifiant (8.27) choisie pour mesurer l'incertitude, non seulement la variété :

$$\vec{\eta} = \overrightarrow{\Upsilon(\vec{\xi})} \qquad (8.29)$$

---

(1) On peut aussi parler de régression d'un groupe de v.a. $Z_1, Z_2, \ldots Z_p$ par rapport au groupe complémentaire $Z_{p+1}, Z_{p+2} \ldots Z_n$.

sera une variété de régression, mais encore les variétés :

$$\overrightarrow{\eta} - \overrightarrow{\eta_0} = \overrightarrow{\gamma\ (\xi)}$$

B) EXEMPLES :

1) Application à la loi de Laplace-Gauss. :

Plaçons-nous dans l'espace à 3 dimensions et supposons que l'ensemble des v.a. Z_1, Z_2, Z_3 obéisse à une loi de Laplace-Gauss et soit :

$$f(z_1, z_2, z_3) = \frac{1}{k} e^{-Q(z_1, z_2, z_3)} \tag{8.30}$$

(où $Q(z_1, z_2, z_3)$ est une certaine forme quadratique définie positive), leur densité de probabilité et considérons l'ellipsoide

$$Q(z_1, z_2, z_3) = 1$$

α) Si $\overrightarrow{y}$ est un vecteur de R_1

Alors quelle que soit l'incertitude ϕ vérifiant (8.3) tout plan conjugué de la direction $0\,z_3$ est une "variété de régression".

β) Si $\overrightarrow{y}$ est un vecteur de R_2

Alors quelle que soit l'incertitude ϕ vérifiant (8.3) toute droite parallèle à la direction conjuguée du plan $0z_2z_3$ est une "variété de régression".

2) Les généralisations de la notion de ligne de régression des moyennes :

α) Si $\overrightarrow{Y}$ est un vecteur de R_1

Alors on obtiendra la "variété de régression des moyennes" en prenant toujours la variance comme mesure de l'incertitude. Il est aisé de voir que toute hypersurface parallèle à l'hypersurface :

$$z_{p+1} = \omega\ (z_1, z_2 \ldots z_p) = \int_{-\infty}^{+\infty} \zeta_{p+1}\, d\, F_{z_1, \ldots z_p}(\zeta p+1) \tag{8.31}$$

sera une variété de régression des moyennes.

β) Si $\overrightarrow{Y}$ est un vecteur de R_m

Alors diverses définitions de "la variété de régression des moyennes" sont possibles. Il est en effet possible alors de donner diverses définitions de l'incertitude $\phi\left[F(y_1, \ldots y_m)\right]$

La plus simple consiste à poser :

$$\phi\left[F(y_1, y_2, \ldots y_m)\right] = \sum_i \min_{a_i} \int (\eta_i - a_i)^2 d\, F(\infty, \infty, \ldots \infty, \eta_i, \infty, \ldots \infty) \tag{8.32}$$

c'est à dire à prendre pour ϕ la somme des variances des variables marginales.

Dans ces conditions, considérons dans un espace à p+1 dimensions l'ensemble des v.a. Z_1, Z_2, ..., Z_p et Z_{p+i}; l'hypersurface de régression de

Z_{p+i} en z_1, z_2 ... z_p sera donnée par (8.31) on aura :

$$z_{p+i} = \omega_i \quad (z_1, z_2, \dots z_p) \tag{8.33}$$

Dans ces conditions, il est aisé de voir que si nous prenons (8.32) pour mesure de l'incertitude, quels que soient b_1, b_2 ... b_m, les variétés :

$$\left.\begin{array}{l} z_{p+1} = b_1 + \omega_1 \quad (z_1, z_2, \dots z_p) \\ \vdots \\ z_{p+m} = b_m + \omega_m \quad (z_1, z_2, \dots z_p) \end{array}\right\} \tag{8.34}$$

sont des variétés de régression.

B) GENERALISATION DE LA NOTION DE DROITE DE REGRESSION

Dans le schéma probabiliste considéré précédemment, nous pouvons imposer à ω $(\overrightarrow{x})$ d'être d'une forme analytique simple. Par exemple, nous pouvons imposer aux composantes du vecteur ω d'être des fonctions linéaires des composantes du vecteur $\overrightarrow{x}$.

En langage ordinaire, nous aurons donc :

$$\left.\begin{array}{l} \omega_1 (z_1, \dots z_p) = \sum_{i=1}^{p} a_{1i} \, z_i + b_1 \\ \omega_m (z_1, \dots z_p) = \sum_{i=1}^{p} a_{mi} \, z_i + b_m \end{array}\right\} \tag{8.35}$$

et nous essaierons de déterminer les valeurs a^*_{ji} et b^*_j pour lesquelles le maximum D^*_1 de D est atteint quand $\overrightarrow{\omega}$ est donné par (8.35).

Il est aisé de voir, compte tenu de (8.3) que le maximum de D sera encore obtenu pour les a^*_{ij} et b_j quels que soient les b_j. Aussi quels que soient les b_j, nous appellerons variété linéaire de régression de Y en x la variété :

$$\left.\begin{array}{l} z_{p+1} = a^*_{11} \, z_1 + \dots + a^*_{1p} \, z_p + b_1 \\ \vdots \\ z_n = a^*_{m1} \, z_1 + \quad + a^*_{mp} \, z_p + b_m \end{array}\right\} \tag{8.36}$$

Il est aisé de voir que dans le cas où l'ensemble des v.a. $Z_1 \dots Z_n$ obéit à une loi de Laplace-Gauss à n dimensions, les variétés linéaires de régression coïncident avec les variétés de régression et ceci quel que soit ϕ.

De plus, quelle que soit la loi de probabilité de X Y quand $\overrightarrow{Y}$ est un vecteur de R_1 et quand on prend la variance comme mesure de l'incertitude on retrouve comme variété de régression l'hyperplan ajusté par la méthode des moindres carrés.

6. — GÉNÉRALISATION DE LA NOTION DE CORRELATION

A) INDICES DE CORRELATION GENERALISES :

Nous appellerons toujours indice de corrélation cf. (6.1) la quantité :

$$C = \frac{\Delta \mathcal{J}}{\max \Delta \mathcal{J}} \tag{8.37}$$

où $\Delta \mathfrak{I}$ est défini par (8.16).

Dans le cas où ϕ est définie par (8.32) on obtient une généralisation du carré du rapport de corrélation de Pearson.

B) INDICES DE CORRELATION DURE GENERALISES :

1) Définition :

Comme précédemment, nous considèrerons encore les indices de corrélation dure :

$$C_{D^*} = \frac{D^*}{\text{Max.}\,\Delta \mathfrak{I}} \qquad (8.37)$$

$$C_{D_1^*} = \frac{D_1^*}{\text{Max.}\,\Delta \mathfrak{I}} \qquad (8.38)$$

qui nous indiquent la valeur des 2 schémas probabilistes les plus intéressants dans la pratique.

Nous nous bornerons ici à donner quelques exemples dans les cas où $\vec{Y}$ est un vecteur aléatoire de R_1 ou de R_2.

2) Exemples :

α) $\vec{Y}$ est un vecteur aléatoire de R_1

Dans ces conditions, si on prend la variance pour mesure de l'incertitude, on a encore $C = C_D^*$ et on retrouve le rapport de corrélation de Pearson.

Par contre, on aura :

$$C_{D_1^*} = \frac{\sigma^2_{z_{p+1}} - \text{var.}\,(Z_{p+1\,.\,12\ldots p})}{\sigma^2_{z_{p+1}}} \qquad (8.39)$$

en désignant comme d'habitude par $z_{p+1\,.\,12\ldots p}$ la v.a.

$$Z_{p+1\,.\,1\;2\ldots p} = Z_{p+1} - \overline{Z}_{p+1} - a^*_{p+1,1} Z_1 - \ldots - a^*_{p+1,p} Z_p \qquad (8.40)$$

et on retrouve aisément le carré du coefficient de corrélation multiple :

$$C_{D_1^*} = R^2 \qquad (8.41)$$

β) $\vec{Y}$ est un vecteur aléatoire de R_2

Un calcul simple montre alors que si on prend pour mesure de l'incertitude la quantité $-\mathcal{D}_1^*$ où $\mathcal{D}_1^*$ est défini par (8.8) on retrouve pour $C_{D_1^*}$ le carré du coefficient de corrélation partielle.

LECTURES RECOMMANDEES

La définition des indices classiques de corrélation se trouve dans Kendall [1] et Brambilla [1]

CHAPITRE IX

ÉTUDE APPROFONDIE DE CERTAINES CLASSES D'INDICES DE CORRÉLATION

Dans les chapitres précédents, nous avons généralisé les notions d'écarts typiques, de gain d'information et de régression et donné une définition nouvelle de la notion de corrélation. Nous ne nous sommes jusqu'à présent occupés que de caractéristiques fonctionnelles théoriques et avons délibérément laissé de côté les questions relatives à la convergence stochastique de suites de caractéristiques fonctionnelles empiriques. Ce problème n'est pas difficile et le lecteur verra sans peine que le statisticien peut en général, à l'aide de données expérimentales, se faire une bonne idée de la valeur de l'indice de corrélation théorique. Nous verrons toutefois qu'alors que la valeur d'un indice de corrélation déduit de l'expérience a une signification simple dans notre théorie, il ne semble pas en être de même des indices de corrélation obtenus à l'aide des définitions classiques.

Avant de pousser la théorie plus avant, remarquons qu'il y a lieu de distinguer entre plusieurs types d'indices de corrélation empiriques.

La valeur de certains indices de corrélation C que nous appellerons indices du type I ou indices qualitatifs, ne dépend que de la probabilité qu'ont X et Y d'appartenir à un nombre fini d'ensembles $\xi_1, \xi_2, \dots, \xi_c$ et $\eta_1, \eta_2, \dots \eta_\ell$. C ne dépendra alors que des probabilités p_{ij} d'observer les couples ξ_i, η_j :

$$C = \Psi \ (p_{ij}) \qquad (9.1)$$

D'autres indices que nous appellerons indices du type III ou indices quantitatifs sont au contraire définis quand X et Y sont des nombres susceptibles de prendre toutes les valeurs x, y entre $-\infty$ et $+\infty$. On a alors :

$$C = \Psi \left[F(x, y) \right] \qquad (9.2)$$

Enfin, d'autres indices que nous appellerons indices du type II sont définis si X appartient à un nombre fini d'ensembles alors que Y est susceptible de prendre toutes les valeurs entre $-\infty$ et $+\infty$.

Nous nous bornerons ici à l'étude des indices des types I et II qui sont les plus simples.

1. — ÉTUDE PARTICULIÈRE DES INDICES DU TYPE I

Nous nous limiterons ici à l'étude de ceux de ces indices C qui sont tels que :

1°/ C = 0 si et seulement si X et Y sont indépendantes

2°/ C = 1 si et seulement si pour i donné tous les p_{ij} sauf un sont égaux à zéro.

Nous avons déjà précédemment rencontré 2 tels indices ; ce sont :

- l'indice de connexion de Gini :

$$\beta_x = \frac{\sum_i \sum_j \left| p_{ij} - p_{i} \cdot p_{.j} \right|}{2 \left[1 - \sum p_{.j}^2 \right]} \tag{9.3}$$

et l'indice quadratique de connexion partielle de Gini :

$$C_x^2 = \frac{\sum_i \sum_j \frac{1}{p_{i.}} (p_{ij} - p_{i.} p_{.j})^2}{\sum_j p_{.j} (1 - p_{.j})} \tag{9.4}$$

A) INDICES THEORIQUES ET INDICES EMPIRIQUES :

Dans la pratique, on ne peut connaître exactement la valeur de l'indice théorique $C = \Psi\ (p_{ij})$ car on ne connait pas les valeurs des p_{ij} .

On connait seulement les fréquences $f_{ij}^{(n)}$ (fréquences d'observation du couple $x_i\ y_j$ au cours de n épreuves indépendantes).

Aussi calculerons-nous l'indice empirique :

$$C^{(n)} = \Psi\ (f_{ij}^{(n)}) \tag{9.5}$$

Ceci posé, nous démontrerons que :

Les indices empiriques de connexion partielle de Gini tendent presque certainement vers leur valeur théorique.

En effet, d'après le théorème de Borel-Cantelli, les $f_{ij}^{(n)}$ convergent presque certainement vers les p_{ij} et $\beta_x^{(n)}$ et $c_x^{2(n)}$ étant des fonctions continues $\Psi\ (f_{ij}^{(n)})$ des $f_{ij}^{(n)}$ ces indices convergent presque certainement vers les $\Psi\ (p_{ij})$ correspondants (1).

B) EXTENSION AU CAS OU X ET Y SONT DES VARIABLES ALEATOIRES CONTINUES :

1) Définition des indices de corrélation théorique et empiriques.

Considérons maintenant un couple aléatoire X Y dont la fonction de répartition F(x,y) est absolument continue.

(1) On peut même affirmer que si C ≠ 0 ou 1 alors la distribution de $C^{(n)}$ tend quand n tend vers l'infini vers une distribution normale (Cf Von Mises (1) ou Hoeffding (1)) - Si C = 1 on a évidemment toujours $C^{(n)} = 1$. Pour le cas où C = 0 voir Von Mises (2) ou (3)

Pour associer un indice du type 1 au couple X Y, il nous faudra effectuer une certaine division en classes. Nous considèrerons c - 1 valeurs particulières de x: $x_1, x_2, \ldots x_{c-1}$ et $\ell - 1$ valeurs de y: $y_1, \ldots y_j, \ldots y_{\ell-1}$ et désignerons par p_{ij} la quantité :

$$p_{ij} = F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}) \quad (9.6)$$

La formule (9.6) permettra également de définir : p_{1j}, p_{i1}, p_{cj}, $p_{i\ell}$ si nous posons symboliquement :

$$x_o = y_o = -\infty \qquad x_c = y_\ell = +\infty$$

L'indice théorique défini sur les p_{ij} donnée par (9.6) dépendra évidemment de la manière dont ont été choisis les vecteurs $\vec{x}$ $(x_1, \ldots x_i, \ldots x_{c-1})$ et $\vec{y}$ $(y_1, \ldots y_j, \ldots y_{\ell-1})$; aussi l'appellerons-nous $C_{\vec{x}\,\vec{y}}$. Si $C_{\vec{x}\,\vec{y}}$ est une fonction continue des p_{ij}, il est clair d'après le lemme précédent que nous pourrons obtenir une approximation $C^{(n)}_{\vec{x}\,\vec{y}}$ de $C_{\vec{x}\,\vec{y}}$ qui tend presque certainement vers $C_{\vec{x}\,\vec{y}}$ quand $n \to \infty$.

Le seul problème théorique qui subsiste est donc de savoir quels renseignements sur F(x, y) nous apporte la connaissance de $C_{\vec{x}\,\vec{y}}$.

Nous nous bornerons d'abord à examiner ce que deviennent les conditions de Fréchet.

2) Propriétés fondamentales de $C_{\vec{x}\,\vec{y}}$

Nous avons les théorèmes suivants :

Théorème I -

Si X et Y sont indépendantes, on a $C_{\vec{x}\,\vec{y}} = 0$

En effet, si X et Y sont indépendantes, on a :

$$F(x, \infty)\, F(\infty, y) = F(x, y) \quad (9.7)$$

relation qui entraine :

$$\left[F(x_i, \infty) - F(x_{i-1}, \infty)\right]\left[F(\infty, y_j) - F(\infty, y_{j-1})\right] = F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}) \quad (9.8)$$

et la relation (9.8) peut s'écrire : $p_{i.}\, p_{.j} = p_{ij}$ (9.9)

(9.9) entraine $C_{\vec{x}\,\vec{y}} = 0$.

Théorème 2.

Si $C_{\vec{x}\,\vec{y}} = 0$, on a pour $x_{i-1} < x < x_i$ et $y_{j-1} < y < y_j$

$$\left| F(x,y) - F(x,\infty)\, F(\infty, y) \right| \leqslant F(\infty, y_j) F(x_i, \infty) - F(\infty, y_{j-1}) F(x_{i-1}, \infty)$$

En effet, on a alors :

$$\left.\begin{aligned} F(x_i, y_j) &= F(x_i, \infty)\, F(\infty, y_j) \\ F(x_{i-1}, y_{j-1}) &= F(x_{i-1}, \infty)\, F(\infty, y_{j-1}) \end{aligned}\right\} \quad (9.10)$$

et en tenant compte de :

$$F(x_{i-1}, y_{j-1}) \leqslant F(x,y) \leqslant F(x_i, y_j)$$
$$F(x_{i-1}, \infty) \leqslant F(x, \infty) \leqslant F(x_i, \infty)$$
$$F(\infty, y_{j-1}) \leqslant F(\infty, y) \leqslant F(\infty, y_j)$$

$$F(x_{i-1}, y_{j-1}) - F(x_i, \infty) F(\infty, y_j) \leqslant F(x,y) - F(x,\infty) F(\infty, y) \leqslant$$
$$F(x_i, y_j) - F(x_{i-1}, \infty) F(\infty, y_{j-1})$$

ou en tenant compte de (9.10)

$$F(x_{i-1},\infty) F(\infty, y_{j-1}) - F(x_i,\infty) F(\infty, y_j) \leqslant F(x,y) - F(x,\infty) F(\infty, y)$$
$$\leqslant F(x_i,\infty) F(\infty, y_j) - F(x_{i-1},\infty) F(\infty, y_{j-1})$$

Remarque :

On peut choisir x_1 y_1 assez petits et x_{c-1} $y_{\ell-1}$ assez grands pour que $F(x_1, \infty) < \eta$ $F(\infty, y_1) < \eta$ $1 - F(x_{c-1}, \infty) < \eta$ $1 - F(y_{\ell-1}, \infty) < \eta$
où η est arbitraire ; alors (quand ℓ et c tendent vers l'infini) on peut trouver un nombre ε tel que :
$F(x_i) - F(x_{i-1}) < \varepsilon$ $F(y_j) - F(y_{j-1}) < \varepsilon$
pour tout i et j entraine :

$$\left| F(x,y) - F(x,\infty) F(\infty, y) \right| < \eta \qquad (9.11)$$

En effet, on a d'après le théorème 2 :

$$\left| F(x,y) - F(x,\infty) F(\infty, y) \right| \leqslant \left[F(x_{i-1},\infty) + \varepsilon \right] \left[F(\infty, y_{j-1}) + \varepsilon \right] - F(x_{i-1},\infty) F(\infty, y_{j-1})$$
$$\leqslant 2\varepsilon + \varepsilon^2$$

quand $x_1 < x < x_{c-1}$ $y_1 < y < y_{\ell-1}$
et $\left| F(x,y) - F(x,\infty) F(\infty, y) \right| < \eta$ dans le cas contraire.

Théorème 3

Si Y est une fonction univalente de x $C_{\vec{x}\,\vec{y}}$ n'est plus en général égal à 1.

En effet, supposons que la ligne de régression de Y en x, $y = \varphi(x)$ coupe la droite $y = y_j$ au point x^* tel que, par exemple : $x_{i-1} < x^* < x_i$
si $\Pr(x_{i-1} < X < x^*) \neq 0$ et $\Pr(x^* < X < x_i) \neq 0$
alors il y aura pour i donné au moins deux p_{ij} différents de zéro et $C_{\vec{x}\,\vec{y}}$ sera différent de 1.

Théorème 4.

Si $C_{\vec{x}\,\vec{y}} = 1$ et si la variable aléatoire Y reste comprise entre les valeurs y_0 et y_1 alors pour tout x on peut trouver deux valeurs de y : y'_x et y''_x telles que :
$y''_x - y'_x \leqslant \text{Max}\,(y_j - y_{j-1})$ et un système admissible de fonctions de répartition liées $F_x(y)$ tel que $F_x(y''_x) - F_x(y'_x) = 1$

En effet, si C = 1 à toute valeur de i correspond une valeur de j telle que $p_{ij} - p_{i\cdot} = 0$

Donc à tout x_i correspond un y_j tel que :

$$F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}) - \left[F(x_i,\infty) - F(x_{i-1},\infty) \right] = 0 \quad (9.12)$$

or :

$$F(x,y) = \int_{-\infty}^{x} F_\xi(y)\, dF(\xi, \infty) \qquad (9.13)$$

et (9.12) s'écrit en tenant compte de (9.13) :

$$\int_{x_{i-1}}^{x_i} F_\xi(y_j)\, dF(\xi,\infty) - \int_{x_{i-1}}^{x_i} F_\xi(y_{j-1})\, dF(\xi,\infty) - \int_{x_{i-1}}^{x_i} dF(\xi,\infty) = 0$$

$$\int_{x_{i-1}}^{x_i} \left\{ F_\xi\ (y_j) - F_\xi\ (y_{j-1}) - 1 \right\} dF(\xi, \infty) = 0 \qquad (9.14)$$

Or, on a : $0 \leqslant F_\xi\ (y_{j-1}) \leqslant F_\xi\ (y_j) \leqslant 1$ la quantité entre parenthèses dans l'équation (9.14) ne peut donc être que négative ou nulle. Elle doit donc s'annuler sauf peut-être sur un ensemble E pour lequel :

$$\int_E dF(\xi, \infty) = 0$$

3) Retour à la notion de fonction univalente

La démonstration précédente nous laisse entrevoir l'impossibilité de vérifier statistiquement que Y est fonction univalente de x. Il est clair en effet que quel que soit le nombre de points expérimentaux, nous pourrons toujours par ces points, faire passer une courbe $y = \varphi(x)$ telle que y soit fonction univalente de x.

En conséquence, si après avoir fait n expériences indépendantes pour déterminer la fonction de répartition du couple aléatoire X, Y, il observe un nuage de points, le statisticien se refusera à considérer comme plausible l'hypothèse que Y est fonction univalente de x au sens de l'analyse et ceci bien que (sauf pour des cas tout à fait exceptionnels) à chaque valeur x_i de X observé correspond une seule valeur de y. S'il agit ainsi, c'est qu'il postule que si Y est fonction univalente de x :

$$y = \varphi(x)$$

a) $\varphi(x)$ doit être fonction continue de x

b) $\varphi(x)$ doit en outre vérifier certaines conditions de régularité que nous allons préciser tout à l'heure.

1° De la liaison bilatérale :

Le plus souvent d'ailleurs, le statisticien suppose qu'il y a liaison bilatérale continue entre X et Y (à toute valeur de x correspond une seule valeur de Y et réciproquement à toute valeur y correspond une seule valeur de X).

S'il en est ainsi, nous pourrons énoncer le théorème suivant :

Théorème :

S'il existe une liaison bilatérale continue entre le couple de v.a. X, Y, de fonction de répartition F(x, y), quel que soit l'ensemble des valeurs de X :

$$x_1 < x_2 < \ldots < x_{c-1}$$

telles que les probabilités marginales $p_{i.} = F(x_i, \infty) - F(x_{i-1}, \infty)$ restent toutes inférieures à ε et quel que soit l'ensemble de valeurs de Y : $y_1 < y_2 < \ldots < y_{\ell-1}$ telles que les $p_{.j} = F(\infty, y_j) - F(\infty, y_{j-1})$ restent compris entre $\lambda\varepsilon$ et ε $(0 < \lambda \leqslant 1)$, le nombre des $p_{ij} = F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1})$ non nuls dans chaque colonne sera au plus égal à $\frac{1}{\lambda} + 2$ et ceci quel que soit ε

Démonstration

Soit $F_x(y)$ la fonction de répartition de Y liée à x. Si Y est fonction univalente de x, nous avons :

$$\begin{aligned} F_x(y) &= 0 \quad \text{si} \quad y < y_0(x) \\ &= 1 \quad \text{si} \quad y \geqslant y_0(x) \end{aligned}$$

et la liaison sera continue si $y_o(x) = \varphi(x)$ est une fonction continue de x.

Posons : $\Psi \quad (x, x + \Delta x, y) = F(x + \Delta x, y) - F(x, y)$ (9.15)

$$= \int_x^{x+\Delta x} F_\xi(y)\, d F(\xi, \infty) \qquad (9.16)$$

a) Si $\varphi(x)$ est une fonction croissante de x, on a :

si $y < \varphi(x)$ $\quad F_\xi(y) = 0$ $\quad$ pour $x < \xi < x + \Delta x$

donc $\Psi \quad (x, x + \Delta x, y) = 0$

si $y > \varphi(x + \Delta x)$, $F_\xi(y) = 1$ $\quad$ pour $x < \xi < x + \Delta x$

donc $\Psi \quad (x, x + \Delta x, y) = F(x, + \Delta x, \infty) - F(x, \infty)$

enfin si $\varphi(x) < y < \varphi(x + \Delta x), F_\xi(y) = 0$ pour $\xi > x + \Delta x$

$$F(x + \Delta x, y) = \int_{-\infty}^{x+\Delta x} F_\xi(y)\, d F(\xi, \infty)$$

$$= \int_{-\infty}^{+\infty} F_\xi(y)\, d F(\xi, \infty)$$

$$= F(\infty, y)$$

$$F(x, y) = F(x, \infty)$$

donc, d'après (9.15) $\Psi \ (x, x + \Delta x, y) = F(\infty, y) - F(x, \infty)$
donc puisque F (y) est une fonction croissante, en vertu de (9.16) $\Psi \ (x, x + \Delta x, y)$ ne pourra varier que pour les valeurs de y telles que :

$$F(x, \infty) \leq F(\infty, y) \leq F(x + \Delta x, \infty)$$

b) si $\varphi(x)$ est une fonction décroissante de x.

Si $\quad y < \varphi(x + \Delta x)$, $F_\xi(y) = 0$ $\quad$ pour $x < \xi < x + \Delta x$

$\Psi \ (x, x + \Delta x, y) = 0$

Si $\quad y > \varphi(x)$ $\quad F_\xi(y) = 1$ $\quad$ pour $x < \xi < x + \Delta x$

$\Psi \ (x, x + \Delta x, y) = F(x + \Delta x, \infty) - F(x, \infty)$

Si $\quad \varphi(x + \Delta x) < y < \varphi(x)$

$$F(x + \Delta x, y) = F(\infty, y) - \int_{x+\Delta x}^{\infty} F_\xi(y)\, d F(\xi, \infty)$$

$$= F(\infty, y) - \left[1 - F(x + \Delta x, \infty)\right]$$

$$F(x, y) = 0$$

et d'après (9.15) $\Psi \ (x, x + \Delta x, y) = F(\infty, y) - \left[1 - F(x + \Delta x, \infty)\right]$
donc $\Psi \ (x, x + \Delta x, y)$ ne pourra croitre que pour un ensemble de valeurs y telles que :

$$1 - F(x + \Delta x, \infty) \leq F(\infty, y) \leq 1 - F(x, \infty)$$

Bref, dès que nous avons entre X et Y une liaison bilatérale continue, il existera deux nombres y' et y'' tels que :

$\Psi \ (x, x + \Delta x, y)$ ne puisse varier que pour $y' \leq y \leq y''$

et que : $F(\infty, y'') - F(\infty, y') = F(x+\Delta x, \infty) - F(x, \infty)$ (9.17)

si donc on pose $x = x_{i-1}$ $x + \Delta x = x_i$

et si y est le plus grand des y inférieurs à y', on aura nécessairement

$$y_{k+1} > y'$$

et $F(\infty, y_{k+1}) + \varepsilon > F(\infty, y'')$

mais $F(\infty, y_j) - F(\infty, y_{j-1}) \geqslant \lambda\varepsilon$ par hypothèse

donc si m est le plus petit entier supérieur à $\frac{1}{\lambda}$

$$F(\infty, y_{k+1+m}) > F(\infty, y'')$$

ainsi au plus $\frac{1}{\lambda}$ + 2 des p_{ij} seront non nuls.

Nous prendrons la propriété précédente comme condition de régularité à imposer aux liaisons unilatérales.

2° De la liaison unilatérale.

Définition :

$F(x, \infty)$ étant supposée absolument continue, nous dirons que nous avons une liaison unilatérale du type $\mathcal{A}$ si pour tout $\varepsilon < \varepsilon_0$ et pour tout ensemble de valeurs de X : $x_1 < x_2 \ldots$ x_{c-1}
telles que $p_{i.} = F(x_i, \infty) - F(x_{i-1}, \infty) < \varepsilon$
et pour tout ensemble de valeurs de Y : $y_1 < y_2 < \ldots y_{\ell-1}$
telles que $\lambda\varepsilon < p_{.j} < \varepsilon$ avec $0 < \lambda \leqslant 1$
le nombre de p_{ij} non nuls pour i donné soit au plus égal à une constante donnée h et ceci quel que soit ε_0

CAS OU ℓ ET c TENDENT VERS L'INFINI :

Revenons au théorème 3, et plaçons-nous pour fixer les idées dans le cas où le couple X, Y admet une densité de probabilité.

Nous savons que si Y est fonction univalente de x, $C^{(n)}_{\vec{x}\,\vec{y}}$ n'est plus en général égal à 1. Nous nous proposons maintenant d'examiner.

a) ce que deviennent les indices classiques quand ℓ et c tendent tous deux vers l'infini de telle sorte que $p_{i.} < \varepsilon$ et $\lambda\varepsilon < p_{.j} < \varepsilon$

b) ce que signifie le fait que les indices classiques sont voisins de 1 quand ℓ et c tendent tous deux vers l'infini.

a/ Examinons d'abord un cas particulier : celui où toute la masse est concentrée sur la droite $y = k.x$ et où les v.a. X et Y ont une densité de probabilité constante sur des segments de longueurs a et k.a.

$$f(x) = 0 \quad \text{si} \quad x \leqslant 0 \text{ ou } x > a$$
$$= \frac{1}{a} \quad \text{si} \quad 0 < x \leqslant a$$

partageons le segment 0, a en c segments égaux à a/c et le segment 0, ka en k.c segments égaux a/c et calculons les valeurs de C_x et C_x^2.

Un calcul simple montre que l'on a alors :

$$\beta_x = \frac{1 - \frac{1}{c}}{1 - \frac{1}{kc}}$$

et par conséquent β_x tendra bien vers 1 quand $c \longrightarrow \infty$

Il n'en sera pas de même de C_x^2 en effet, on a :

$$C_x^2 = \frac{c \left(\frac{1}{kc} - \frac{1}{kc^2}\right)^2 \quad c\,k + c\left(\frac{1}{k\,c^2}\right) \quad (k\,c^2 - kc)}{1 - \frac{1}{kc}} \tag{9.19}$$

et C_x^2 tendra vers $\frac{1}{k}$ quand $c \longrightarrow \infty$

Cet exemple prouve qu'il est toujours dangereux de considérer sans étude préalable les indices de corrélation (même s'ils vérifient les conditions de Fréchet) comme un outil pour déterminer s'il y a une relation fonctionnelle entre X et Y.

Aussi ne nous semble-t-il pas que, pour le moment du moins, on puisse donner au mot indice de corrélation un autre sens précis que celui que nous avons donné au chapitre VI.

Pas plus que l'indice C_x^2, l'indice de Jordan (1) qui vaut :

$$J^2 = \frac{1}{\ell - 1}\left[-1 + \sum \sum \frac{p_{ij}^2}{p_{i.}\,p_{.j}}\right] \tag{9.20}$$

(et par conséquent n'est pas un indice de corrélation à notre sens) ne semble pouvoir servir à isoler les cas où il y a une liaison fonctionnelle entre X et Y. En effet, on a dans le cas précédent :

$$J^2 = \frac{1 - \frac{1}{c}}{k\left(1 - \frac{1}{c\,k}\right)} \tag{9.21}$$

et par conséquent J^2 tendra vers $\frac{1}{k}$ quand c tend vers l'infini.

Il est vrai que Mme Geiringer [1] ayant remarqué pour J^2 un inconvénient de la nature que nous venons de signaler, propose de le remplacer par l'indice :

$$g^2 = \frac{\ell\, J^2}{1 + (\ell - 1)\, J^2}$$

Cet indice n'est pas non plus un indice de corrélation à notre sens. Il tend évidemment vers 1 dans l'exemple sus-mentionné. Nous verrons toutefois qu'il ne semble pas non plus nous permettre de juger s'il y a une relation fonctionnelle entre X et Y.

Montrons que les résultats auxquels nous sommes parvenus sont généraux. Autrement dit, montrons que si nous avons une liaison unilatérale du type $\mathcal{A}$ et si ε est de l'ordre de grandeur de $\frac{1}{c}$, alors que J^2 reste, en général, inférieur à 1, g^2 et β_x tendent vers 1.

a) J^2 reste en général inférieur à 1.

En effet, si pour tous les p_{ij}, on a $p_{ij} < \mu p_{i.}$, on a en posant pour simplifier :

$$S = \sum_i \sum_j \frac{p_{ij}^2}{p_{i.}\,p_{.j}} \tag{9.23}$$

(1) Pour les propriétés de cet indice, voir Fréchet (2)

d'où on tire :

$$S < \mu \sum_i \sum_j \frac{p_{ij}}{p_{.j}}$$

$$S \leqslant \mu \ell$$

donc :

$$J^2 \leqslant \frac{\mu \ell - 1}{\ell - 1}$$

et le second membre tend vers μ quand $\ell \longrightarrow \infty$

b) g^2 tend vers 1.

En effet, comme dans chaque colonne on a au plus un nombre fini fixe k de $p_{ij} \neq 0$ le plus grand des p_{ij} de chaque colonne sera au moins égal à $\frac{p_{i.}}{k}$. De plus, comme $p_{i.} < \varepsilon$ au moins $\frac{1}{\varepsilon}$ des $p_{i.}$ seront supérieurs à $\frac{1}{c}$ mais $p_{.j} > \lambda \varepsilon$ donc pour au moins $\frac{1}{\varepsilon}$ valeurs de i, on aura :

$$\frac{p_{ij}}{p_{i.}\, p_{.j}} \geqslant \frac{1}{k} \cdot \frac{\frac{1}{ck}}{\lambda \varepsilon}$$

et $S \geqslant \frac{1}{C\, \varepsilon^2 k^2 \lambda}$ Si donc on a pris ε de l'ordre de grandeur de $\frac{1}{c}$, $S \longrightarrow \infty$

et :

$$g^2 = \frac{\ell}{\ell - 1} \; \frac{S-1}{S} \qquad (9.24)$$

tendra vers 1.

c) β tendra vers 1.

En effet, si dans chaque colonne au plus k des p_{ij} sont non nuls, on pourra considérer le numérateur N de l'indice de connexion partielle de Gini comme la somme de deux nombres.

$$N = N' + N''$$

où $N' = \sum{}' \left| p_{ij} - p_{i.}\, p_{.j} \right|$ la somme $\sum{}'$ étant étendue aux seules valeurs de i et j pour lesquelles $p_{ij} \neq 0$ et $N'' = \sum{}'' \left| p_{ij} - p_{i.}\, p_{.j} \right|$ la somme $\sum{}''$ étant étendue aux seules valeurs de i et de j pour lesquelles $p_{ij} = 0$; dès lors on aura $\sum{}' p_{i.}\, p_{.j} + \sum{}' p_{ij} \geqslant N' \geqslant \sum{}' p_{ij} - \sum{}' p_{i.}\, p_{.j}$

or :

$$\sum p_{ij} = \sum{}' p_{ij} = 1$$

tandis que $\sum p_{i.}\, p_{.j} \leqslant c\, k\, \varepsilon^2$ donc tend vers zéro

et N' tend vers 1.

On a de même :

$$N'' = \sum{}'' p_{i.}\, p_{.j}$$

$$= \sum p_{i.}\, p_{.j} - \sum{}' p_{i.}\, p_{.j}$$

$$1 \geqslant N'' \geqslant 1 - c\, k\, \varepsilon^2$$

donc $N \longrightarrow 2$

le dénominateur D tend lui aussi vers 2, en effet, on a

$$\mu \varepsilon < p_{.j} < \varepsilon$$

mais $\sum_j p_{.j} = 1$

donc $$\ell \mu \varepsilon < 1 < \ell \varepsilon$$

$$\ell \varepsilon < \frac{1}{\mu}$$

Dès lors $\sum p_{.j}^2 < \ell \varepsilon^2$ tend vers 0 et $D \longrightarrow 2$

donc $$\beta_x \longrightarrow 1 \qquad (9.25)$$

Réciproquement

On peut se demander ce que signifie le fait d'observer que $C > 1 - \eta$

Nous examinerons successivement ce qui se passe pour l'indice de Jordan, l'indice de Mme Geiringer et l'indice simple de connexion partielle de Gini.

a) L'indice de Jordan :

La démonstration précédente prouve qu'alors le plus grand des p_{ij} devra être au moins égal à $(1-\eta)\, p_i$.

Mais il est aisé de voir qu'au moins $\ell(1-\eta)$ des rapports $\lambda_{ij} = \frac{p_{ij}}{p_{i.}}$ doivent être au moins égaux à $1-\eta$

En effet on a :

$$S = (\ell - 1) J^2 + 1 \qquad \text{donc } S > \ell\,(1-\eta)$$

et de l'autre :

$$S = \sum_i \sum_j \lambda_{ij} \frac{p_{ij}}{p_{.j}}$$

or $$\ell_{.j} = \sum_i \lambda_{ij} \frac{p_{ij}}{p_{.j}} \leqslant 1$$ donc au moins $\ell(1-\eta)$ des sommes $\ell_{.j}$ devront être supérieures à $1-\eta$ et ceci n'est possible que si un au moins des λ_{ij} de chacune d'elles est supérieur à $1-\eta$

Il est à remarquer que pour $\eta < \frac{1}{2}$ on ne peut avoir deux $\lambda_{ij} > 1-\eta$ dans la même colonne, donc :

Théorème :

Si $J^2 > 1-\eta$ $\left(\eta < \frac{1}{2}\right)$ il existe un nombre j tel que $p_{ij} > (1-\eta)\, p_{i.}$ pour au moins $\ell(1-\eta)$ valeurs de i

b) L'indice de Madame Geiringer :

La démonstration précédente a montré qu'il suffisait que S tende vers l'infini pour que l'indice de Mme Geiringer tende vers 1. Pour qu'il en soit ainsi, il suffira évidemment que quand le nombre des lignes et des colonnes augmente indéfiniment, pour une infinité de valeurs de p_{ij}, les rapports $\frac{p_{ij}}{p_{i.}}$ et $\frac{p_{ij}}{p_{.j}}$ restent supérieurs à un nombre fixe ε.

Il en sera ainsi notamment si une masse finie est tout entière concentrée sur une ligne, le reste de la masse étant réparti d'une manière quelconque.

c) L'indice de connexion partielle de Gini :

Si $\beta > 1-\eta$ pour $\lambda \varepsilon < p_{.j} < \varepsilon$

alors :

$$\sum p_{.j}^2 < \ell \varepsilon^2 < \frac{\varepsilon}{\lambda}$$

et le numérateur devra être supérieur à :

$$2\,(1-\eta)\,(1-\frac{\varepsilon}{\lambda}) = 2\,(1-\eta') \text{ où } \eta' \text{ est petit.}$$

Soit α_{ij} le plus grand des 2 nombres p_{ij} ou $p_{i.}\,p_{.j}$ et μ_{ij} le plus petit. On aura :

$$N = \sum_i \sum_j \alpha_{ij} - \sum_i \sum_j \mu_{ij}$$

or :
$$\sum_i \sum_j p_{ij} = 1 \qquad \sum_i \sum_j p_{i.}\,p_{.j} = 1$$

donc :
$$\sum_i \sum_j \alpha_{ij} + \sum_i \sum_j \mu_{ij} = 2$$

d'où :
$$\sum_i \sum_j \mu_{ij} < \eta'$$

On a donc à définir 2 ensembles complémentaires E et E' tels que :

$$\sum_E p_{ij} = \varepsilon_1 \qquad \sum_{E'} p_{i.}\,p_{.j} = \varepsilon_2$$

avec :
$$\varepsilon_1 + \varepsilon_2 \leqslant \eta'$$

Ainsi le fait que $\beta \longrightarrow 1$ quand $\varepsilon \longrightarrow 0$ ne signifie pas que Y est fonction univalente de x. Il en sera par exemple encore ainsi si toute la masse est concentrée sur un nombre fini de lignes.

2. — ÉTUDE PARTICULIÈRE DES INDICES DU TYPE II

A) INDICES THEORIQUES ET INDICES EMPIRIQUES :

Nous avons vu que, pour les indices du type II, X n'est susceptible d'appartenir qu'à un nombre fini c d'ensembles $\xi_1, \xi_2, \ldots \xi_c$ avec les probabilités $p_1, p_2, \ldots p_c$, alors que Y est susceptible de prendre toutes les valeurs de $-\infty$ à $+\infty$. Soit $F_{\xi_i}(y)$ la fonction de répartition de la variable liée Y_{ξ_i}. C sera alors fonction uniquement des p_i et $F_{\xi_i}(y)$;

nous noterons $\quad \mathscr{C} = \Lambda\left[\, p_i \,,\, F_{\xi_i}(y) \,\right]$

Nous nous bornerons ici à l'étude des indices du type II qui vérifient les conditions suivantes :

1°/ C = 0 si et seulement si :

$$F_{\xi_1}(y) = F_{\xi_2}(y) = \ldots = F_{\xi_c}(y)$$

2°/ C = 1 si et seulement s'il correspond à tout ξ_i un nombre $y_0(\xi_i)$ tel que :

$$F_{\xi_i}(y) = 0 \quad \text{si} \quad y < y_0(\xi_i)$$
$$F_{\xi_i}(y) = 1 \quad \text{si} \quad y \geqslant y_0(\xi_i) \qquad (9.26)$$

Les indices du type II seront dits continus si à tout nombre η on peut faire correspondre un nombre ε tel que :

$$\left| p_i - p_i' \right| < \varepsilon$$

et borne sup. $\left| F_{\xi_i}(y) - F'_{\xi_i}(y) \right| < \varepsilon \quad (i = 1, 2 \ldots c)$ (9.27)

entrainent :

$$\left| C\left[p'_i, F'_{\xi_i}(y)\right] - C\left[p_i, F_{\xi_i}(y)\right] \right| < \eta \tag{9.28}$$

Ceci posé, si $f_i^{(n)}$ et $F_{\xi}^{(n)}(y)$ désignent les fréquences de ξ_i et les fonctions de répartition empiriques des variables liées Y_{ξ_i} observées au cours de n épreuves indépendantes, on sait d'après les théorèmes de Borel et Glivenko-Cantelli que quand $n \longrightarrow \infty$ les quantités $\left| f_i^{(n)} - p_i \right|$ et borne sup. $\left| F_{\xi_i}^{(n)} - F_{\xi_i}(y) \right|$ tendent presque certainement vers zéro.

Donc :

lemme. - Les indices empiriques continus du type II tendent presque certainement vers leur valeur théorique quand $n \longrightarrow \infty$

Regardons donc si l'indice simple de connexion de Gini :

$$\mathcal{G} = \frac{\sum_i p_i \int_{-\infty}^{+\infty} \left| F_{\xi_i}(y) - F(\infty, y) \right| dy}{2 \int_{-\infty}^{+\infty} F(\infty, y)\left[1 - F(\infty, y)\right] dy} \tag{9.29}$$

est un indice continu.

Cet indice n'est évidemment défini que si les intégrales du numérateur et du dénominateur convergent.

Il en sera ainsi notamment si l'on peut trouver 2 nombres y' et y'' tels que $F(\infty, y') = 0 \quad F(\infty, y'') = 1$ (9.30)

Ceci posé, pour pouvoir démontrer que $\mathcal{G}$ est un indice continu, nous serons obligés de supposer que (9.30) est vérifiée. S'il en est ainsi, les conditions (9.27) entraineront :

$$\left| \sum_i p_i F'_{\xi_i}(y) - \sum p_i F_{\xi_i}(y) \right| < \varepsilon$$

$$\left| F'(\infty, y) - F(\infty, y) \right| < \varepsilon$$

$$\int \left| \left\{ \left| F_{\xi_i}(y) - F(\infty, y) \right| - \left| F'_{\xi_i}(y) - F'(\infty, y) \right| \right\} \right| dy \leq 2\varepsilon\,(y'' - y')$$

de même la variation du dénominateur tend vers zéro et, si Y ne se réduit pas presque certainement à une constante, on aura $\mathcal{G}' - \mathcal{G} \longrightarrow 0$, bref, l'indice simple de connexion de Gini, empirique tendra presque certainement vers sa valeur théorique si l'étendue de Y est bornée et non nulle.

B) EXTENSION AU CAS OU X ET Y SONT DES V.A. CONTINUES :

La démonstration p (9. 9) prouve que si y est fonction univalente monotone de x

$$y = \varphi(x)$$

et si on prend des points de subdivision tels que :

$$F(x_i, \infty) - F(x_{i-1}, \infty) < \varepsilon$$

alors on pourra trouver pour chaque colonne 2 nombres y' et y'' tels que la probabilité pour qu'un nombre soit de la ième colonne et reste inférieur à y :

$$F_i(y) = F(x_i, y) - F(x_{i-1}, y)$$

ne soit susceptible de varier qu'entre y'_i et y''_i

$$F_i(y'_i) = 0$$

$$F_i(y''_i) = F_i(\infty)$$

et que : $F(\infty, y''_i) - F(\infty, y'_i) < \varepsilon$

Définition -

$F(x, \infty)$ étant supposée absolument continue, nous dirons que nous avons une liaison unilatérale du type $\mathcal{A}$ si pour tout $\varepsilon < \varepsilon_0$ et à tout ensemble de valeurs de x

$x_1 < x_2 < \ldots < x_c$ telles que $F(x_i, \infty) - F(x_{i-1}, \infty) < \varepsilon$ on peut faire correspondre une suite de nombres y'_i y''_i tels que :

$$F_i(y''_i) - F_i(y'_i) = F_i(\infty)$$

et que : $F(\infty, y''_i) - F(\infty, y'_i) < k\varepsilon$ (où k est une constante donnée).

Examinons quelle valeur prend l'indice de connexion simple de Gini quand nous avons une liaison unilatérale du type $\mathcal{A}$ et quand :

$$F(x_i, \infty) - F(x_{i-1}, \infty) < \varepsilon$$

posons : $F_{\xi_i}(y) = \dfrac{F_i(y)}{F_i(\infty)}$

Le numérateur de l'indice simple de connexion de Gini est :

$$N = \sum p_i \int_{-\infty}^{+\infty} \left| F_{\xi_i}(y) - F(\infty, y) \right| dy \qquad (9.31)$$

Or, il existe par hypothèse, pour chaque valeur de ξ_i un nombre $y_0(\xi_i)$ tel que si nous définissons y'_i et y''_i par :

$$F(\infty, y'_i) = F(\infty, y_0) - \frac{k\varepsilon}{2} \qquad F(\infty, y'') = F(\infty, y_0) + \frac{k\varepsilon}{2}$$

on aura : $F_{\xi_i}(y'_i) = 0 \qquad F_{\xi_i}(y''_i) = 1$

pour y_1 fixé on a : $\left| F_{\xi_i}(y_1) - F(\infty, y_1) \right| = F(\infty, y_1)$

pour ceux des ξ_i tels que $y'_i > y_1$ donc avec une probabilité au moins égale à :

$$1 - F(y'_1) = 1 - F(y_1) - \frac{k\varepsilon}{2}$$

de même, on aura $F_{\xi_i}(y_1) - F(\infty, y_1) = 1 - F(\infty, y_1)$ pour ceux des ξ_i tels que $y''_i < y_1$ donc avec une probabilité au moins égale à $F(y) - \frac{k\varepsilon}{2}$

Bref, le numérateur vaudra au moins :

$$N = \int \left\{ F(\infty, y)\left[1 - F(\infty, y) - \frac{k\varepsilon}{2}\right] + \left[1 - F(\infty, y)\right]\left[F(\infty, y) - \frac{k\varepsilon}{2}\right] \right\} dy$$

et si l'étendue de la v.a. marginale Y est finie non nulle, l'indice simple de connexion de Gini tend bien vers 1.

LECTURES RECOMMANDEES

I. Sur les conditions de Fréchet et indices vérifiant ces conditions.

Consulter Fréchet [2], Féron [1], Brambilla [1], Jordan [3], Geiringer [1], Gini [1], Salvemini [1].

II. Sur les indices peu connus vérifiant d'autres conditions que celles de Fréchet.

Darmois [3] ou Risser et Traynard [1] ou Tchuprov [1], Fréchet [2] ou Féron [1], Brambilla [1], Gini [2] [3] [4] [5], Karl Pearson [2] [3] [4] [5], Steffensen [1], Kendall [2]

III. Sur la loi de probabilité de diverses caractéristiques fonctionnelles empiriques.

A/ dans le cas de petits échantillons.

1) borne supérieure de la variance des caractéristiques fonctionnelles empiriques :

Fréchet [1], Darmois [5], Cramér [1]

2) Le calcul effectif de la distribution est évidemment un cas d'espèce. On lira avec fruit sur toutes ces questions :

Kendall [1] et Insee [1], Fisher [1] et en général tous les articles de E.S. Pearson.

B/ dans le cas des grands échantillons. On a alors des théorèmes qui nous permettent d'affirmer la tendance vers la loi normale sauf dans des cas exceptionnels : cf Von Mises [1], Hoeffding [1] Bernstein [2] [3]
Même dans les cas exceptionnels on peut d'ailleurs souvent déterminer la distribution limite Von Mises [2] [3]

Donnons pourtant quelques indications complémentaires :

1. pour les divers indices de corrélation :

a) pour le coefficient de corrélation linéaire cf. E.S. Pearson [8] et [10], Pittman [2], Chesire, Oldis et Pearson [1], David [1]

b) pour le rapport de corrélation de Pearson cf Fisher p. 605 et Hotelling [1]

c) pour les coefficients de corrélation du rang cf Kendall [2]

2. pour les autres caractéristiques fonctionnelles :

a) écarts typiques : Pour l'étendue, E.S. Pearson [4] [12] [14], pour l'écart type Kendall [1], Davies et E.S. Pearson [1], pour l'écart moyen E.S. Pearson [15]

b) Différence moyenne Nair [1]

c) pour les coefficients β_1, β_2, w_n E.S. Pearson [7] [9] [13]

d) pour le coefficient de variation E.S. Pearson [11]

Si l'on veut utiliser une caractéristique fonctionnelle récente, on sera généralement amené à déterminer sa distribution approximative. On y parviendra en suivant l'une des méthodes qui ont déjà été utilisées pour la détermination des caractéristiques fonctionnelles citées ci-dessus. Il s'agit là en général d'un travail simple mais laborieux.

CHAPITRE X

GÉNÉRALISATION DE CERTAINS TYPES DE PROCESSUS STOCHASTIQUES

Nous vous proposons maintenant de généraliser les notions de processus à v.a. indépendantes, de chaine simple de Markow, de processus stochastiques stationnaires et de processus à accroissements indépendants.

1. — PROCESSUS A VARIABLES ALÉATOIRES INDÉPENDANTES

Un tel processus n'est guère concevable que dans le cas discret.

On a une suite de v.a. $X_1, \ldots, X_n \ldots$.
et l'ensemble des n premières v.a. a pour fonction de répartition $F(t_1 \ldots t_n)$ Soit $X_n \mid 1 \ldots n-1$ la v.a. de fonction de répartition.

$$F(t_n \mid x_1, \ldots x_{n-1}) = \frac{F(x_1 \ldots x_{n-1}, t_n)}{F(x_1 \ldots x_{n-1}, \infty)} \text{ posons } \mathcal{F}(t_n) = F(\infty, \ldots, \infty, t_n)$$

A) SENS CLASSIQUE :

On aura un processus à v.a. indépendantes au sens classique si

$$F(t_n \mid x_1, \ldots, x_{n-1}) = \mathcal{F}(t_n) \qquad (10.1)$$

et ceci quels que soient $x_1 \ldots x_n$ (sauf peut-être pour un ensemble de valeurs de $x_1 \ldots x_n$ de probabilité nulle).

Donc au sens classique on aura un processus à v.a. indépendantes si quel que soit n fini les vecteurs aléatoires $\vec{\mathfrak{X}}(X_1, \ldots X_{n-1})$ et $\vec{\mathfrak{Y}}(X_n)$ forment un couple aléatoire de vecteurs indépendants.

Et nous avons vu que si Φ est une fonctionnelle strictement concave la condition nécessaire et suffisante pour qu'il en soit ainsi est que

$$\Delta \mathfrak{I} \mathfrak{Y} \mid \mathfrak{X} = 0 \qquad (10.2)$$

B) PREMIERE GENERALISATION - PROCESSUS A V.A. INDEPENDANTES RELATIVEMENT FONCTIONNELLE Φ (Sens strict)

Si nous n'imposons plus à Φ d'être une fonctionnelle strictement concave mais simplement d'être une fonctionnelle concave alors la relation (10.2) n'entraine plus (10.1) nécessairement. Aussi toutes les fois que (10.2) sera vérifiée dirons-nous que nous avons un processus à v. a. indépendantes relativement à la fonctionnelle Φ au sens strict.

On notera en passant que si l'on prend la variance comme mesure de l'incertitude on tombe sur la très importante classe des processus à v. a. indépendantes en moyenne et dont la variance est finie.

DEUXIEME GENERALISATION - PROCESSUS A V.A. INDEPENDANTES RELATIVEMENT A LA FONCTIONNELLE Φ (Sens large)

Nous obtiendrons des conditions encore moins restrictives si au lieu d'imposer à $\Delta \mathcal{J}_{\vec{\mathcal{Y}}\ \vec{\mathcal{X}}}$ d'être nul nous imposons seulement à une des mesures de l'information dure d'être nulle.

Dans la pratique nous n'aurons que deux cas à considérer suivant que l'on prend D^* ou D_1^* pour mesure de l'information dure.

1) PROCESSUS A V.A. INDEPENDANTES AU SENS LARGE PREMIER GENRE.

Nous dirons que nous avons un tel processus si quel que soit $n < \infty$

$$D^*_{\vec{\mathcal{Y}}\,\vec{\mathcal{X}}} = 0 \tag{10.3}$$

Nous avons vu qu'il ne peut en être ainsi que si dans le schéma probabiliste considéré au chapitre 6 le couple de vecteurs aléatoire $\vec{\mathcal{X}}\ \vec{\mathcal{Y}}$ est remplacé par un couple de vecteurs aléatoires indépendants $\vec{\mathcal{X}}^*\ \vec{\mathcal{Y}}^*$ dans le schéma probabiliste le mieux adapté.

Dans le cas particulier où l'on prend la variance pour mesure de l'incertitude on retrouve les processus formés de v.a. (à écart type borné) et indépendantes en moyenne.

2) PROCESSUS A V.A. INDEPENDANTES AU SENS LARGE 2e GENRE:

Nous dirons que nous avons un tel processus si

$$D_1^* = 0 \tag{10.4}$$

Nous avons vu qu'il en est ainsi si lorsqu'on impose à la nouvelle ligne de régression de $\mathcal{Y}^*$ en $\mathcal{X}^*$ d'être un hyperplan, le couple de vecteurs aléatoires $\mathcal{X}\mathcal{Y}$ est dans le schéma le mieux adapté remplacé par le couple de vecteurs aléatoires indépendants $\vec{\mathcal{X}}^*\ \vec{\mathcal{Y}}^*$

En particulier si on prend la variance comme mesure de l'incertitude un processus à v.a. indépendantes au sens large du 2e genre est un processus à variables aléatoires non carrelés

$$E(x_s\, x_t) = E(x_s)\, E(x_t)$$

2. — CHAINES SIMPLES DE MARKOV

Considérons un espace abstrait Ω de points ω et un certain corps de Borel $\mathcal{F}$ - Autrement dit on suppose qu'on a :

1) $\Omega \in \mathcal{F}$

2) si $\Lambda \in \mathcal{F}$ alors $\Omega - \Lambda \in \mathcal{F}$

3) $\bigcup_1^n \Lambda_i \in \mathcal{F}$, $\bigcap_1^n \Lambda_i \in \mathcal{F}$

On suppose qu'on associe à ces ensembles une probabilité $P(\Lambda)$ c'est-à-dire une fonction non négative complètement additive et telle que $P(\Omega) = 1$

On suppose en outre qu'à tout nombre réel t de l'ensemble T on associe une fonction B - mesurable $x_t(\omega)$

A) SENS CLASSIQUE :

On a un processus de Markov au sens classique si pour tout entier $n > 1$ et quels que soient les nombres $t_1 < t_2 \ldots < t_n$ on a pour tout Λ

$$P\left\{x_{t_n}(\omega) \leqslant \lambda \mid x_{t_1}, x_{t_2} \ldots x_{t_{n-1}}\right\} = P\left\{x_{t_n}(\omega) \mid x_{t_{n-1}}(\omega)\right\} \qquad (10.5)$$

Il est aisé de voir que si Φ est une fonctionnelle strictement concave il faut et il suffit que quelque soit l'entier $n < \infty$ et les t_i choisis

$$\Delta \mathfrak{J}\ X_{t_n}\ X_{t_{n-1}} = \Delta \mathfrak{J}\ X_{t_n}\ \vec{\mathcal{X}} \qquad (10.6)$$

$\vec{\mathcal{X}}$ représente le vecteur aléatoire $x_{t_1} \ldots x_{t_{n-1}}$

pour qu'on ait un processus de Markov au sens classique.

B) SENS STRICT :

Aussi dirons-nous comme précédemment que nous avons un processus de Markov au sens strict relativement à la fonctionnelle Φ si quel que soit l'entier n (10.6) est vérifiée.

Si en particulier nous prenons la variance comme mesure de l'incertitude, nous aurons un processus de Markov par rapport à la variance au sens strict si quel que soit n

$$E\,(X_{t_n} \mid x_{t_{n-1}}) = E\,(X_{t_n} \mid x_{t_1} \ldots x_{t_{n-1}}) \qquad (10.7)$$

<u>Remarque</u> :

Dans le cas particulier où $E(X_{t_n} \mid x_{t_{n-1}}) = x_{t_{n-1}}$ ce dernier processus devient une martingale.

C) SENS LARGE :

De même que précédemment nous dirons que nous avons un processus de Markov du premier genre au sens large si quel que soit n

$$D^*_{X_{t_n}},\ x_{t_{n-1}} = D^*_{X_{t_n}},\ \vec{\mathcal{X}} \qquad (10.8)$$

Il est aisé de voir que si nous prenons la variance pour mesure de l'incertitude nous aurons un tel processus si et seulement si (10.7) est vérifiée.

De même nous aurons un processus de Markov du second genre au sens large si le D^*_1 relatif au couple $X_{t_{n-1}}, X_{t_n}$ est le même que le D^*_1 relatif aux vecteurs $\vec{\mathcal{X}}\,(X_{t_1} \ldots X_{t_{n-1}})$ et $\vec{\mathcal{Y}}\,(X_{t_n})$

En particulier si l'on prend la variance comme mesure de l'incertitude on aura un tel processus si $E(\mathcal{Y}^* \mid \mathcal{X}^*)$ a la même valeur quand $\vec{\mathcal{X}}$ représente le seul vecteur $X_{t_{n-1}}$ ou le vecteur $X_{t_1} \ldots X_{t_{n-1}}$ - En d'autres termes si la variance étant prise comme mesure de l'incertitude $\hat{E}\,(X_{t_n} \mid x_{t\ n-1})$ représente la meilleure approximation linéaire de X_t. On doit avoir :

$$\hat{E}\,(X_{t_n} \mid x_{t_1} \ldots x_{t_{n-1}}) = \hat{E}\,(X_{t_n} \mid x_{t_{n-1}}) \qquad (10.9)$$

3. — PROCESSUS STOCHASTIQUES STATIONNAIRES

A) SENS CLASSIQUE :

C'est un processus tel que quels que soient $t_i \in T$ $(i = 1, 2, \ldots n)$ et la constante $h \in \mathcal{H}$ où $\mathcal{H}$ est un ensemble tel que la distribution de $X_{t_1+h} \ldots X_{t_n+h}\left(F_h(\lambda_1, \ldots \lambda_n)\right)$ soit indépendante de h.

(1) Un tel processus a déjà été considéré par Doob [1] p. 90.

On peut exprimer ce fait d'une autre manière - Remarquons que quelles que soient les valeurs $h_1 \ldots h_m$ de $H \in \mathcal{X}$ on peut construire un couple de vecteurs aléatoires $\overrightarrow{\mathcal{H}}$ (H), $\overrightarrow{\mathcal{X}}(X^{(1)}, \ldots X^{(n)})$

$$1^\circ \ Pr(H = hj) = pj > 0 \quad \Sigma pi = 1$$

$$2^\circ \ Pr(X_i < \lambda_i \mid H = hj) = F_{h_j}(\lambda_1 \ldots \lambda_n)$$

Dans ces conditions si le processus stochastique est stationnaire au sens classique alors

$$\Delta \mathfrak{I}_{\vec{\mathcal{X}}, \vec{\mathcal{H}}} = 0 \qquad (10.10)$$

B) SENS STRICT :

Aussi dirons-nous que nous avons un processus stochastique station - naire au sens strict relativement à la fonctionnelle concave Φ, si quels que soient les vecteurs aléatoires $\vec{\mathcal{H}}$ et $\vec{\mathcal{X}}$ choisis (10.10) est vérifiée -

Il est aisé de voir que si Φ est une fonctionnelle strictement concave, tout processus stationnaire au sens strict est stationnaire également au sens classique. Il n'en est pas de même si Φ est seulement une fonctionnelle concave.

4. — PROCESSUS A ACCROISSEMENTS INDÉPENDANTS

A) SENS CLASSIQUE :

On a un tel processus si pour $t_1 < t_2 \ldots < t_n$ $n \geqslant 3$ les v.a.

$$\mathcal{Y} = x_{t_2} - x_{t_1} \text{ et } Z = x_{t_n} - x_{t_{n-1}}$$

sont indépendantes.

Il en résulte que si nous avons un tel processus nous aurons toujours quels que soient $t_1 \ldots t_n$

$$\Delta \mathfrak{I}_{Z\mathcal{Y}} = 0 \qquad (13.13)$$

B) SENS LARGE :

<u>1er genre</u> - On aura un tel processus si $D^*_{Z\mathcal{Y}} = 0$

<u>2e genre</u> - On aura un tel processus si $(D^*_1)_{Z\mathcal{Y}} = 0$

Dans le cas où la variance est prise comme mesure de l'incertitude, il en sera évidemment ainsi si on a affaire à un processus à accroissements orthogonaux.

LECTURES RECOMMANDEES

Doob [1] , Blanc Lapierre et Fortet [1]

BIBLIOGRAPHIE

BERNSTEIN S.

[1] Fondements géométriques de la théorie des corrélations Metron 1928

[2] Sur les sommes de grandeurs aléatoires liées de classes AN et BN C.R. Ac. Sc. URSS 32 p. 303-8 ; 1941.

[3] Sur l'extension du théorème limite des probabilités aux sommes de quantités dépendantes.
Math. Ann. 97, 1927, 1-59.

BLANC LAPIERRE ET FORTET

[1] Théorie des fonctions aléatoires.

BOURBAKI

[1] Utilisation des nombres réels en topologie générale L III Ch. IX.

BRAMBILLA

[1] Teoria statistica della correlazione e della connessione (Mariano Ricci Florence 1942).

BRAVAIS

[1] Analyse mathématique sur les probabilités des erreurs de situation d'un point.
C.R. t. 9, 1846 p. 255

CANTELLI

[1] Sulla differenza media con ripetizione
Giorn. Econ. e Rev. di Stat 1913.

L. CHESIRE, E. OLDIS, E.S. PEARSON

[1] Further experiments on the sampling distribution of the correlation coefficient.
J. Amer. Statist. Ass. 27 n° 178 121-8

COCHRAN ET COX

Experimental Designs Wiley 1950.

CRAMER H.

[1] Mathematical methods of Statistics (Princeton)

[2] Random Variables and Probability Distributions Cambridge tracts in mathematics n° 36 Cambridge 1937.

CURTISS

[1] On the distribution of the quotient of two chance variables Ann. Math. Statist. 12, 4, p. 409-22, Déc. 1941.

DARMOIS G.

[1] Liaison de probabilité - Analyse factorielle XVIIIe Congrès Intern. phil. Sc.

[2] Conférence faite à l'Institut Henri-Poincaré en 1949.

[3] Statistique mathématique Doin 1928.

[4] La statistique et ses applications A. Colin.

[5] Leçons sur l'estimation statistique. Imprimerie Nationale 1947.

[6] Répartitions à deux caractères. Liaisons entre deux grandeurs. Barometro economico 6/11 p. 647 (1934).

[7] Sur les régressions linéaires et paraboliques Ass. Fr. av. Sc. Congrès de la Victoire t.II 1945 p. 121-3

DAVID (F.N.)

[1] Tables of the ordonates and Probability integral of the distribution of the correlation coefficient in small samples Biometrika publications.

DIEUDONNE

Sur le théorème de Lebesgue Nlkodym

[1] Part. I Ann. of. Math. t. 42, 1941, p.547-55
[2] Part. II Bull. Soc. Math. Fr. t. 72 1944 p. 193-239
[3] Part. III Ann. Univ. Grenoble 1948 p. 25-48

O DAVIES et E.S.PEARSON

Methods of estimating samples the population standard deviation J.R. Sup. I, 76-93, 1934.

DOOB

[1] Stochastic Processes, 1953.

[2] Stochastic processes with an integral valued parameter. Tr. Amer. Math. Soc. 44, 1938 p. 386-92.

EISENHART HASTAY, WALLIS.

Selected techniques of statistical analysis Mc Graw Hill 1947.

FELLER

[1] An introduction to probability theory and its applications.

FERON

[1] Mérites comparés des divers indices de corrélation
Journ. Soc. Statist. Paris, sept-oct. 1947 - p.328-52.

[2] De l'information C.R. t. 230 p.1495-7.

[3] Tests d'hypothèse pour les tableaux à 2 lignes et 2 colonnes (conférence au séminaire de calcul des probabilités).

[4] Information et corrélation C.R. t. 234 1343-5.

[5] Convexité et information C.R. t. 234 1840-1.

[6] Lois asymptotiques du calcul des probabilités
Formulaire de Math. pour Phys. et Ing. CNRS, fasc. 12 Calcul des Probabilités 1952.

[7] De la régression C.R. t.234 p. 2143-5.

[8] Nouvelles généralisations des notions de moyenne et de variance. Applications aux processus stochastiques. C.R. du séminaire de calcul des probabilités.

FERON ET FOURGEAUD

[1] Information et régression - C.R. t.232 p. 1636-8.

DE FINETTI

[1] Sui metodi proposti per il calcolo della differenza media
Metron 9 1931.

DE FINETTI ET PIACIELLO

[1] Calcolo della differenza media
Metron 8 n° 3, 89, 1930.

FISHER

[1] Contributions to mathematical statistics Wiley 1950.

[2] Statistical methods for research workers Ohver et Boyd.

[3] The goodness of fit of regression formulae and the distribution of regression coefficients JRSS 85, p.597, 1922.

FORTET

[1] Eléments de calcul des probabilités. CNRS 1950.

FRECHET

[1] Sur les fonctionnelles continues Ann. Ec. Norm. 1910 t. 27.

[2] Anciens et nouveaux indices de corrélation - leur application au calcul des retards économiques. Econom. Janv. 1947 p. 30

[3] Nouvelles définitions de la valeur moyenne et des valeurs équi-probables d'un nombre aléatoire.
Ann. de l'Univ. de Lyon Sect. A IX p.5-26, 1946.

[4] Le coefficient de connexion statistique de Gini-Salvemini.
Mathematica 23, p. 46-51 - 1947-1948.

[5] Sur un essai infondé pour sauver le coefficient classique dit de corrélation
Rev. Inst. Intern. Statist. 1950 p.314.

[6] Sur les tableaux de corrélation dont les marges sont données.
Ann. Univ. Lyon 9 , 1951 p.53-77.

[7] Sur une nouvelle définition des positions typiques d'un élément aléatoire abstrait.
Ann. Ec. Norm. Sup. C.R. 226, 1948, p. 1420-21.

[8] Une propriété générale des valeurs typiques d'un nombre aléatoire
Publ. ISUP VI, f 1, 1952.

[9] Généralités sur les probabilités - variables aléatoires -
Gauthier Villars 1937.

[10] Sur le coefficient dit de corrélation et sur la corrélation en général.
Rev. Inst. Int. St 1 1933 p. 1-8.

[11] Rapport de la commission d'étude de l'usage du coefficient de corrélation.
R. Inst. Intern. Stat. 4 2 p. 238-42 , 1936.

[12] Etudes Scientifiques d'ordre statistique sur le coefficient de linéarité dit de corrélation.

[13] Les indices statistiques de dépendance fonctionnelle.
Bar. Econ. 6, 4, p. 231-2, 1934.

[14] Sur l'extension de certaines évaluations statistiques au cas de petits échantillons.
Rev. Inst. Intern. Statist. 1943 p. 182.

[15] Les valeurs typiques d'ordre nul ou infini d'un nombre aléatoire.
Rev. Inst. Intern. Stat. 1948 p. 1-22.

[16] Les espaces abstraits

Mme GEIRINGER

[1] Korrelations-messung auf Grund der Summenfunktion
Zeitsch angew Math. Mech. 13, p. 123, 1933.

[2] Korrelationsmodelle Zeitsch angew Math. Mech. 1933 p. 19-35.

GINI

[1] Di una misura della dissomiglianza tra due gruppi di quantita e delle sue applicazioni allo studio delle relazioni statistiche.
Atti del R. Ist. veneto di Sc Let ed Arti 1914 t. 74.

[2] Indici di omofilia e di rossomiglianza e loro relazione col cœfficiente di correlazione e con gli indice di Attrazione même périodique t. 74.

[3] Sul criterio di concordanza tra due caretteri, même périodique t. 75.

[4] Indici di concordanza, même périodique t. 75.

[5] Sur le coefficient de corrélation comme caractéristique du degré de concordance R. Inst. Int. Statist. 4 355-66 1936.

GUILBAUD

[1] La Cybernétique (Que sais-je ? 1954).

GRUZEWSKA

[1] Sur la distribuante de deux variables aléatoires dépendantes. C.R. 19 Nov. 1951 p. 1256.

[2] Un schéma probabiliste de processus stochastiques. C.R. 26 Nov. 1951 p. 1345-6

HALMOS

[1] Measure Theory.

HARDY. LITTLEWOOD, POLYA.

[1] Inequalities (Cambridge Univ. Press. 1934).

HARTLEY. R.V.L.

[1] Transmission of information Bell System Tech. J. 8 535 1928.

HELMERT

[1] Uber die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler Zeit. für Math. and Phys. 20 300.

HILLE

[1] Functional analysis and semi groups Ann. Math. Soc. coll. Publ. 1948

HOEFFDING

[1] A class of statistics with asymptotically normal distribution Ann. Math. Statist. 19, 3, p. 293-325 sept. 1948.

HOEFFDING ET ROBBINS

[1] The central limit theorem for dependent random variables Duke Math. J. 15, 3 sept. 1948.

HOTELLING

[1] The distribution of correlation ratios calculated from random data. Proc. Nat. Akad Sc. VII 657-62 1925

INSEE

[1] Bibliographie des principales tables statistiques.

JACKSON

[1] Note on the median of a set of numbers. Bull. Ann. Soc. Math. 27 p. 160, 1921.

[2] Note on quartiles and allied measures Bull. Am. Math. Soc. 29 1923 p. 17-20.

JORDAN (Charles)

[1] Statistique mathématique (voir en particulier p. 172)

[2] Les coefficients d'intensité relative de Korosy, J. Soc. Hongr. Stat. 1927

[3] Critique de la corrélation au point de vue de la probabilité. Hermann 1938.

[4] Sur une série de polynômes dont chaque somme partielle représente la meilleure approximation d'un degré donné suivant la méthode des moindres carrés.
Proc. London Math. Soc. 1921.

Mc KAY ET E.S. PEARSON

[1] A note on the distribution of range in samples of n.
Biom. 25 415-20, 1933.

M.G. KENDALL

[1] The advanced theory of Statistics London 1943.

[2] Rank correlation Griffin 1948.

LAPLACE (Marquis de)

[1] Théorie analytique des probabilités 1818.

LEBESGUE

[1] Leçons sur l'intégration et la recherche de fonctions primitives Gauthier-Villars 1928.

LEVY P.

[1] Théorie de l'addition des variables aléatoires.

[2] Sur la notion de probabilité conditionnelle.
Bull. Sc. Math. 60, 1936, 66-71.

MARKOV

[1] Wahrscheinlichkeitsrechnung 1912.

R. Von MISES

[1] Les lois de probabilité pour les fonctions statistiques.
Ann. Inst. H.P. 6, 1936, 185-212.

[2] On the asymptotic distribution of differentiable statistical functions.
Ann. Math. Statist. 18, 309-48, 1947.

[3] Théorie et application des fonctions statistiques.
Univ Roma Rendic Mat. 1952 11 n° 3-4 374-410

MOOD

[1] Introduction to the theory of statistics.

NAIR

[1] The standard error of Gini's mean difference
Biom. 28, 428-36, 1936.

NARUMI

[1] On the general forms of bivariate frequency distributions which are mathematically possible when regression and variation are subjected to hiviting conditions.
Biom. 15 p. 77 et 209 (1923).

J. NEYMAN ET E.S.PEARSON

[1] On the Use and interpretation of certain test criteria for purposes of statistical inference.
Biom. 20 A p. 175-240 et 263-294.

[2] On the problem of the most efficient tests of statistical hypotheses
Phil.Trans.231, 289-337 (1933).

[3] The testing of statistical hypotheses in relation to probabilities a priori.
Proc. Cambr. Phil. Soc. 29, 492-510, 1933.

[4] Contribution to the theory of testing statistical hypotheses
Stat. Res. Memoirs I, 1, 37, 1936.

[5] Sufficient statistics and uniformely most powerful tests of statistical hypotheses.
Stat. Res. Memoirs I,1-37.

PARENTI

[1] Estensione di una relazione di Frechet al coefficiente parabolica di ordine.
Gior. Econom. Mai Juin 1941 p. 319-24.

E.S.PEARSON

[1] On Polychoric coefficients of correlation.
Biom. 14, 127-56, 1922.

[2] The probable error of a class-index correlation.
Biom. 14 261-80, 1923.

[3] Note of the approximations to the probable error of a coefficient of correlation.
Biom. 16, 196-8, 1924.

[4] A further note on the distribution of range in samples taken from a normal population.
Biom. 18, 173-94, 1926.

[5] Further note on the linear correlation ratio.
Biom. 19, 223-4, 1927.

[6] Some notes on sampling tests with two variables.
Biom. 21, 337-60, 1929.

[7] A further development of tests of normality.
Biom. 22, 239-49, 1930.

[8] The test of significance for the correlation coefficient
J. Amer. Statist. Ass. 26, n° 174, 128-34, 1931.

[9] Note on tests for normality.
Biom. 22, 423-4, 1931.

[10] The test of signifiance for correlation coefficient some further results.
J. Amer. Statist. Ass. 27, n° 180, p.424-6, 1932.

[11] Distribution of the coefficient of variation. Comparison of K.T Mc Kay's approximation with experimental sampling results.
J. Roy. Statist.Soc. 55, 703-4 (1932).

[12] The percentage limits for the distribution of range in samples from a normal population (n=100)
Biom.24, 404-17.

[13] A comparaison of β_2 and Mr Geary's w_n criteria
Biom.27 333-52, 1935

[14] The probability integral of the range in samples of n observations from a normal population.
Foreword to Tables. Biom.32, 301-8,1942.

[15] The probability integral of the mean deviation; editorial note.
Biom.33, 252-3, 1945.

[16] The choice of statistical tests illustrated on the interpretation of data classed in a 2 × 2 table.
Biom. 24, 139-167, 1947.

PEARSON K.

[1] On the theory of contingency and its relation to association and normal correlation Draper's company research memoirs.
Biom. Series 1904.

[2] On a coefficient of class heterogeneity or divergence.
Biom. vol. 5 1901 p. 198.

[3] On the general theory of skew correlation and non linear regression Draper's Comp. Research Memoirs, Biom Ser II, 1905.

[4] Note on the history of correlation.
Biom.13, p. 25, 1920.

PITMAN

[1] The estimation of location and scale parameters of a continuous population of any given form.
Biom. 30, 391, 1939.

[2] Significance tests which may be applied to samples of any population II. The correlation coefficient test Sup. JRSS 4,p. 225,1938

POMPILJ

[1] Sulla regressione rend. di mathematica Roma 1946.

PRETORIUS

[1] Skew bivariate frequency surfaces examined in the light of numerical illustrations.
Biom. 22, 1930, p. 109-223.

SAKS

[1] Théorie de l'intégrale (Hafner).

SALVEMINI

[1] Nuovi procedimenti di calcolo degli indici di dissomigliana e di connessione. Statistica 9,3-26 - 1949.

[2] Su la misura della stretezza o rigore di una relazione statistica Statistica 1945-6 292-315.

[3] Independenza statistica tra due o piu variabili.
Atti della IX, X, XI, reunione della societa di statistica.

[4] Sul calcolo degli indici di concordanza tra due caratteri quantitative
Atti della VI Riun Soc. Ital. Stat. 1943.

[5] Sugli indici di omofilia
Att. 1 Riun `Soc. Ital. Stat. oct. 1939.

[6] L'indice di dissomiglianza fra distribuzioni continue
Metron 1951.

[7] Sui vari indici di cograduazione
Statistica 9, 2, 1951.

[8] Dissomiglianza a piu dimensioni.
Atti della IX, X, XI Riun della Soc. di Stat. 1951.

RADO

Subharmonic Functions

RISSER

[1] D'un certain mode de recherche des surfaces de probabilités.
C.R. 225, 1366-68, 1947.

RISSER ET TRAYNARD

[1] Les principes de la statistique mathématique.
Gauthier Villars.

SARMANOV

[1] De la corrélation isogène (en russe).
C.R. Ac.Sc.URSS 1945.

SCHUTZENBERGER

[1] Contribution aux applications statistiques de la théorie de l'information (thèse de doctorat 1954).

SHANNON ET WEAVER.

[1] The mathematical Theory of communication (Univ. Illinois Press)
reproduction de Bell. Tech. Syst. J. 1948 p. 379-423 et 623-56

SNEDECOR

[1] Statistical methods (Collegiate press).

STEFFENSEN

[1] On certain measures of dependence between statistical variables.
Biom. 1934, Ann. IHP 3, 326, 1933.

TCHUPROV

[1] Grundbegriffe und Grund problem der Korrelations theorie
Leipzig 1932.

USPENSKY

[1] Introduction to mathematical probability New York 1937.

de la VALLEE-POUSSIN

[1] Intégrales de Lebesgue, fonctions d'ensemble, classes de Baire Gauthier-Villars 1934.

VILLE

[1] Sur la convergence de la médiane des n premiers résultats d'une suite infinie d'épreuves indépendantes.
C.R.203, p. 1 309, 1936.

WIENER

[1] Cybernetics or control and communication in the animal and the machine (Hermann 1948)

[2] The extrapolation, intrapolation and smoothing of stationary times series with engineering applications (Wiley 1949).

WISHART

[1] A note on the distribution of the correlation ratio.
Biom.24, 441, 1932.

TABLE DES MATIÈRES

Pages

IV

TABLE DES MATIÈRES DU VOLUME V

ANNÉE 1956

Pages

FASCICULE 1

FASCICULE 2

FASCICULES 3 et 4 REUNIS

J. & R. SENNAC, Imprimeurs
54, Fg Montmartre, PARIS (9e)